Scattering and Mobility in Indium Gallium Arsenide Channel, Pseudomorphic High Electron Mobility Transistors (InGaAs pHEMTs)

John Lawson Pearson

Thesis Submitted to the University of Glasgow for the Degree of Doctor of Philosophy, Containing Research Undertaken in the Department of Physics and Astronomy

March 1999

ABSTRACT

Extensive transport measurements have been completed on deep and shallow-channelled InGaAs p-HEMTs of varying growth temperature, indium content, spacer thickness and doping density, with a view to a thorough characterisation, both in the metallic and the localised regimes. Particular emphasis was given to MBE grown layers, with characteristics applicable for device use, but low measurement temperatures were necessary to resolve the elastic scattering mechanisms. Measurements made in the metallic regime included transport and quantum mobility – the former over a range of temperatures between 1.5K to 300K. Conductivity measurements were also acquired in the strong localisation regime between about 1.5K and 100K.

Experimentally determined parameters were tested for comparison with those predicted by an electrostatic model. Excellent agreement was obtained for carrier density. Other parameters were less well predicted, but the relevant experimental measurements, including linear depletion of the 2DEG, were sensitive to any excess doping above a 'critical' value determined by the model.

At low temperature (1.5K), it was found that in all samples tested, transport mobility was strongly limited at all carrier densities by a large q mechanism, possibly intrinsic to the channel. This was ascribed either to scattering by the long-range potentials arising from the indium concentration fluctuations or fluctuations in the thickness of the channel

layer. This mechanism dominates the transport at low carrier densities for all samples, but at high carrier density, an additional mechanism is significant for samples with the thinnest spacers tested (2.5nm). This is ascribed to direct electron interaction with the states of the donor layer, and produces a characteristic transport mobility peak. At higher carrier densities, past the peak, quantum mobility was found only to increase monotonically in value. Remote ionised impurity scattering while significant, particularly for samples with intermediate (5nm) and thin (2.5nm) spacers, was never found to dominate.

As has been reported for similar structures, anisotropy of transport mobility was found, with the $[01\bar{1}]$ direction having a higher mobility than the [011] direction ((100) GaAs substrate nominally aligned ±0.1°). Intermediate directions had intermediate mobilities. The anisotropy increased with indium content and growth temperature, and persisted to at least 300K. In addition, we found that quantum mobility was independent of direction suggesting that the mechanism responsible is dominated by short-range, large q scattering.

Both transport and quantum mobility were reduced when donor layer correlations were removed using the process of bias cooling. Quantum mobility was more sensitive to this process although excess donors in the doping layer also affected values at high carrier densities. Applying Matthiessen's rule to both correlated and uncorrelated transport mobility data, strongly suggested that remote ionised impurity scattering was consistent with theory for samples with a 5nm spacer, but that an additional mechanism, as mentioned above, must exist in the samples with a 2.5nm spacer.

Variable temperature studies further revealed that at low carrier densities, weak localisation was present, with strong, temperature

dependent, activated transport also apparent at higher depletion. At high carrier densities in the thinnest spacer samples (2.5nm), a transport mobility peak evolved with decreasing temperature. The mechanism responsible was undetermined, but it was reminiscent of weak localisation-like behaviour. (515 words)

TABLE OF CONTENTS

LIST OF FIGURES

energy of the Fermi level elsewhere requires an additional self-consistent calculation because the spatial distribution of charge requires knowledge of the band profile, which in turn demands knowledge of the charge distribution. Note also that the Fermi level lies in the first subband in the channel and hence the electrons trapped here are degenerate - the 2 dimensional electron gas (2DEG).

Figure 1.2.4 The full conduction band-edge energy profile, now extended from fig. 1.2.3 to include the upper layers. The example shown has a 10nm spacer, with 30% indium concentration in the channel, which is 'deep'. The relative energies of the Fermi level and the conduction band-edge require self-consistent calculations, but to a good approximation it may be taken as pinned at 70meV below the conduction band in the Si δ-doped layer. Ordinarily, there is a significant population of electrons trapped in the *DX* or deep donor centres there. The GaAs monolayers, around the δ-layer are not shown. An effective channel width (a) is introduced and an explanation is given in the text below.

Figure 1.3.1 The plan view of the growing GaAs substrate, where the [100] direction of growth is chosen. Either coloured sphere can represent Ga or As, but one layer (colour) lies in a plane below the other. The growing plane of a subsequently grown ternary will look much like this with some Ga atoms replaced with In or Al atoms.

Figure 1.3.2a A schematic representation of the substrate, channel and spacer- each in isolation showing the lattice mis-match. The values of a_2 and a_3 are very similar, with a_1 different. The relative differences in lattice parameter are much exaggerated for visibility. The layer widths shown are, of course, much thinner than in reality.

Figure 1.3.2b The extension to fig. 1.3.2a showing the effect of the pseudomorphic strain on the channel layer in particular. As the substrate is assumed infinitely thick then most of the change will occur in the lattice parameter of the channel, that is, a_1 tends to a_2. To compensate there is a tensile strain applied to the bonds in the vertical direction (Poisson effect). These strained vertical bonds may also influence the bonding within the spacer layer above.

Figure 1.3.3 The formation of a dislocation (D), and the extra plane of atoms (shown bordered). By this re-arrangement of atoms, the lattice

can reduce the potential energy stored in its bonds. These dislocations can however severely degrade device performance.

Figure 1.4.1 The energy relationship between the aluminium mole fraction and the energies of the various *DX* centres and the hydrogenic states.[33] Note, in particular, the comparative values at x=0.3. It should also be remarked that the density of the *DX* centres decreases with both increasing assigned number (0,1,2 and 3) and decreasing x[35].The variation with x of the energy of the *DX*0 centre is known,[36] but it is an assumption that the other lines lie parallel, although the energy difference at some x values is known.

Figure 1.4.2 Schematic 2-D cross-sectional view of a *DX* centre, showing the lattice distortion. The blue circles are group III atoms, the yellow circles are As atoms and the red circle is a donor atom/ion. (after ref.[16])

Figure 1.4.3 A typical configuration-coordinate diagram of the 4 known electronic states of a given donor centre. The blue parabola represents the D^+ state; green- the D^0 state; red- the DX^- state; and purple represents a state used in photo-ionization. See the text for an explanation.

Figure 1.4.4 A schematic representation of the distortion of the conduction band-edge in the vicinity of a *DX* centre.[16] E_{DD} represents the average energy of the two trapped electrons relative to the conduction band-edge (E_C) and, if enough occupied *DX* centres of this type exist, the Fermi level can pin on this blue line. If the Al mole fraction (x) is 0.3, then the effective E_{DD} is about 70meV.

Figure 2.1.1 Conduction band-edge energy profile, See the text for an explanation of the symbols.

Figure 2.4.1 Scattering Diagram in $\boldsymbol{k}$-space – $\boldsymbol{k}$ is the initial wavevector; $\boldsymbol{k}+\boldsymbol{q}$ is the final wavevector and $\boldsymbol{q}$ is the resultant. $Sin(\theta/2)=|\boldsymbol{q}|/2|\boldsymbol{k}|$.

Figure 2.4.2 Schematic energy diagram showing, comparatively, the various scattering potentials of atoms *A*, and *B* and the weighted average of the two, taking relative abundance into account (E_A,E_B and E_{all} respectively). (after Singh[17])

Figure 5.9.12 A similar plot as fig. 5.9.10, but for sample A1236 at a gate voltage of –2.70V. There is a similar degree of fit between experiment and theory as sample A1235 at a gate voltage of –2.75V, see fig. 5.9.11.

Table 5.9.1 A table of experimental and calculated theoretical parameters relevant to a study of weak localisation affected by temperature (T) and magnetic field (B).The samples used were A1235 and A1236. **KEY:**-Vg - gate voltage; T – temperature; μ- transport mobility ($x10^4 cm^{-2}V^{-1}s^{-1}$); n- carrier density ($x10^{12}cm^{-2}$); B_{Tr} – characteristic transport magnetic field; τ_ϕ/τ - ratio of phase to transport relaxation time; $\Delta\sigma$ - conductivity correction for temperature (T) and magnetic field (B) in units ($e^2/\pi h$). DT refers to the difference in conductivity between the two temperatures at a given gate voltage – the numbers taken from the column on the immediate left. DB represents the same procedure but applied to the column on its immediate left. The temperatures shown are those where a measurement could be made from both the temperature data- figs. 5.9.8 and 9 and the magnetic field data – figs. 5.9.11-13. $\sigma(T)$ is the value of the conductivity at a given temperature and gate voltage taken directly from figs. 5.9.8 and 9, and $\Delta\sigma(B)$ is the weak localisation correction at a given temperature based on fitting the experimental curves at a given temperature and gate voltage to the theoretical curves using the ratio τ_ϕ/τ as a fitting parameter (see equation 5.9.2) and B_{Tr} calculated using equation 5.9.1, with n and μ values taken from other experimental data. With the τ_ϕ/τ ratio determined, $\Delta\sigma(B)$ can be calculated using equation 2.5.12.

CHAPTER 1

GENERAL INTRODUCTION AND CRYSTAL THEORY

1.1 General Introduction

1.1.1 Outline

The electronic transport properties associated with the formation of junctions between two or more single-crystalline semiconductor materials have been investigated for many years. This research has been undertaken in both the 'metallic' (high carrier density) and the 'insulating' (hopping) regimes to characterise these systems. An exhaustive review[1] is available of the early examples of these, particularly the silicon-based **m**etal-**o**xide-**s**emiconductor **f**ield-**e**ffect **t**ransistors (MOSFETs).

With the development of molecular beam epitaxy (MBE) from the early 1970s, along with other advanced methods for the epitaxial growth

of III-V compound semiconductors, the production of high quality field-effect transistors (FETs) has accelerated. This is particularly so for high mobility FETs, where the current is directed along the degenerate quasi-two-dimensional electron gas (2D.E.G.) This 'gas' can be formed when specific compounds are successively grown, whose conduction band-edge energy configuration permits the chemical potential (or Fermi energy as it will be known here) to lie above the band-edge in this region. To achieve such a system, high quality epitaxial growth is necessary, incorporating a low defect and impurity density, with the layers separated by abrupt interfaces. Section 1.2 discusses the construction of the potential energy and the electrostatics of the valence and conduction bands generated by epitaxial growth. See table 1.1.1 for the typical layers used in this work.

One of the junction-pair combinations extensively researched and tested is the gallium arsenide/aluminium gallium arsenide ($GaAs/Al_xGa_{1-x}As$) heterojunction. These materials produce a lattice-matched junction over all of x, which is the mole fraction of aluminium in the ternary alloy. This matching avoids lattice strain, whose subsequent relaxation can lead to severe mobility, and hence device, degradation. Varying x offers some flexibility in the conduction band edge offset value, and hence the electron confinement (section 1.2). 2DEG mobilities greater than 10^7 $cm^2V^{-1}s^{-1}$ are presently attainable in devices based on this heterojunction.

Unfortunately, this mobility is only available at temperatures close to absolute zero, where only the effects of impurity and defect scattering are seen. These can be minimised in the high purity, low defect samples grown presently. However, such temperatures are totally impractical for the use of these FETs, and it is found that by

Layer Type	Thickness
Surface	
GaAs Cap	7.5 or 20nm
$Al_{0.3}Ga_{0.7}As$ Upper Spacer	10 or 60nm
GaAs Monolayers	2 monolayers
Si Delta-doping 3.5 or $7x10^{12}cm^{-2}$	
GaAs Monolayers	3 monolayers
$Al_{0.3}Ga_{0.7}As$ Spacer	2.5, 5 or 10nm
$In_{0.3/0.2}Ga_{0.7/0.8}As$ Channel	10nm
GaAs Undoped Buffer	600nm
$Al_{0.3}Ga_{0.7}As$/GaAs 100 Period Superlattice	9 monolayers each
GaAs Substrate (100)	

Table 1.1.1 The general layer systems studied. The substrate is the wafer platform on which single-crystalline, epitaxial growth can proceed. The superlattice prevents the migration of defects and impurities towards the active region above. The buffer helps prevent the Fermi energy pinning on the energy of these trapped charges. The channel confines the active current-producing electrons. The spacer offers a buffer between these electrons and the mobility-reducing ionised donors – designated by the Si delta-doping. This deposition is sandwiched by GaAs monolayers to help prevent migration of the donors. The upper spacer merely provides the main buffer between the gate metal on the surface and the electrons in the channel. The cap provides a suitable surface for the gate deposition.

100K, optical phonon scattering becomes the major mobility limiting factor.[(2)] The mobility falls by many orders of magnitude until room temperature is attained.

As conductivity is proportional to both carrier density and mobility, good conduction values would, ideally, have both parameters as large as possible. Unfortunately, GaAs/AlGaAs based devices have a

relatively low carrier density when compared with other system possibilities. Therefore, they are not ideal for device use. As a consequence, methods of generating a high carrier density focused much of the attention.

It turns out that the key factor in achieving this is the choice of the materials in and around the channel layer. As will be seen later in this chapter, one of the best methods is that exploiting what is called 'pseudomorphic' growth. In this growth mode however, a certain amount of elastic strain must be accommodated when these materials are grown successively. The reward is a much deeper channel energy-well with an associated higher carrier density. There are however, disadvantages as well as advantages in the use of these materials and these will now be discussed.

1.1.2 The Advantages and Disadvantages of Pseudomorphic Growth

It was already known that the ternary compound indium gallium arsenide ($In_yGa_{1-y}As$) offers a lower effective electron mass and a higher electron saturation velocity, when compared with the conventional GaAs transport layer described in section 1.1.1. However, these systems have a further advantage, which is particularly relevant to increasing the channel conductivity. Larger conduction band-edge energy offsets can be produced, which can increase electron confinement if the offset is formed between the channel and spacer layers. These terms will be explained in section 1.2 and illustrated in figures 1.2.(1-4). Indeed, if both these layers are composed of ternary materials ($In_yGa_{1-y}As$ and

$Al_xGa_{1-x}As$ respectively), excellent flexibility in engineering a desired value of offset can be attained. This can be achieved by adjusting the electron affinity and bandgap of the materials by varying the relative atomic proportions, x and y.

In interfacing two layers in this manner, two disadvantages emerge. With, for example, GaAs/AlGaAs or some indium phosphide/indium gallium arsenide (InP/InGaAs) systems, successive epitaxially grown layers are effectively lattice matched, and there is little strain induced in the chemical bonds at an interface. Resulting defects, such as dislocations, are then minimised. With a ternary channel material, such as $In_yGa_{1-y}As$, grown on a GaAs substrate, a large strain effect can be induced, see section 1.3. The amount of strain depends on the magnitude of y and hence the lattice-mismatch. It turns out, however, that if the channel width is less than a certain critical width, the strain can be elastically accommodated in the chemical bonds with minimal dislocation production. This critical thickness is mole fraction dependent,[3] and if exceeded, misfit dislocations will form, severely degrading device performance. Therefore, provided one works within these constraints, a measure of flexibility is possible in designing adequately deep channel wells.

The other perceived disadvantage is unavoidable. In introducing a ternary material in which to accommodate the 2DEG electrons, the problem of intrinsic random alloy disorder scattering of the channel electrons is considerable. Also the alloy atoms should distribute themselves randomly in available sites, but statistically, one of the constituent atom species can segregate then aggregate, initiating a local cluster scattering mechanism.[4] Electron microscopic evidence[5] of such indium clustering or concentration fluctuations can prove useful in

establishing length scales and specifying their dependence on the indium mole fraction.

Considerable effort has been expended in modelling alloy scattering in the metallic regime since one of the initial attempts of Littlejohn et al.[6] The main theoretical problem is the treatment of the alloy scattering potential, which has been assigned many different forms, including the Dirac-delta function of Luo et al,[7] and the shaped discontinuity of the crystal potential of Basu and Nag.[8] Bhattacharyya et al[9] present a theoretical and experimental comparison over a range of temperatures from 1.7K to 300K. The three alloy scattering potential treatments of Littlejohn et al[6] were used. These were the electronegativity (EN) and the electron affinity (EA) difference of the binary end compounds, and the energy band gap difference (EG) of the unalloyed components. All the other major scattering factors were included in the analysis, and although the model using the EN treatment produced mobilities closest to the experimental temperature dependent mobility, none fitted perfectly. This problem is typical and highlights the need for experimental data to fit an alloy scattering potential value or other information for input to the theoretical models.

Therefore, much of the work in this report has been promoted to provide experimental data in the metallic and localised regimes. The former to compare experimental data with theoretical models with regard to alloy scattering etc., and the latter to provide further information regarding the indium concentration fluctuations and perhaps extend that to the metallic regime. Unlike channels composed of binary compounds, ternary varieties, such as InGaAs, have their low temperature mobility strongly limited by alloy or related scattering, and only towards room temperature does optical phonon scattering dominate.

The study of the pseudomorphic layers is conveniently sub-divided into two:

1. From a physical perspective, where the properties of, and the carrier transport within, the chosen materials are analysed.
2. From an engineering perspective, where the device characteristics, for example, current, noise, etc. are optimised for implementation either discretely or as part of an integrated circuit.

The approach followed in this work was primarily from the first perspective, but also with a view to answering questions posed by both device modelling and the practical implementation in 'ultra-fast' circuitry.

Pseudomorphic growth research is not limited to this type of system and many other materials, such as ZnSe on GaAs, AlGaAs or InGaAs have been studied.[10] Likewise, many other layer combinations, such as single, double and multiple quantum wells or superlattice structures[11] have also been analysed.

1.2 Formation of Heterostructures

1.2.1 Introduction

A heterojunction is formed when two crystalline materials of differing band-gap are grown successively, with an abrupt junction between them. They exist in two forms - isotype and anisotype. The former term applies when the two materials have the same doping type and the latter when they contain different doping types. An understanding of the mechanism of formation, which confers many of the unique characteristics of these structures, is crucial to this work.

The existence of the so-called channel and its associated two-dimensional electron gas (2DEG) relies on the manner in which the appropriate layers merge together. The best approach to its understanding is from the perspective of energy, where the effect of the differing band-gaps can be highlighted. As our materials are predominantly n-type, we will concentrate on the construction of the conduction band-edge energy profile.

It should be emphasised from the outset that the relation between the energy of the conduction band-edge and the spatially distributed electronic charge density requires a self-consistent approach. These qualifications will be asserted as they arise, and the formation and properties of the 2DEG will involve a thorough explanation in this and most of the other sections of this chapter.

1.2.2 The Energetics of Heterojunction Formation

The materials used to form the electrically active region around the channel of the heterostructural system are shown in fig. 1.2.1.[12,13] The channel will contain a highly mobile, degenerate density of electrons.

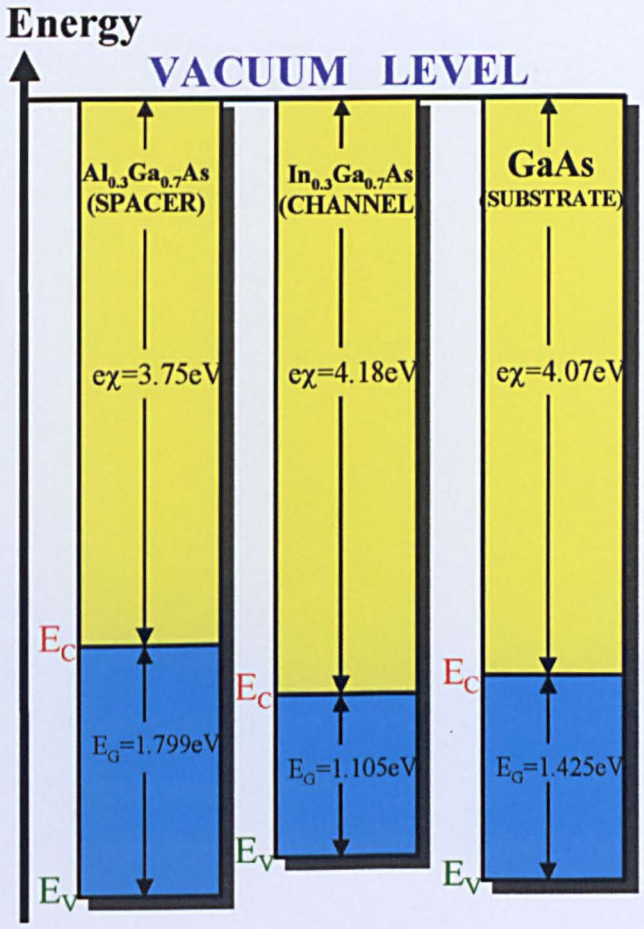


Figure 1.2.1 Schematic energy band diagram of three isolated semiconductors at room temperature, including the vacuum level which can act as a reference. E_C and E_V are the conduction and valence band edges respectively. $e\chi$ represents the electron affinity and E_G is the energy band-gap. Note that there would be an offset in E_C and E_V between the semiconductors if they were joined. See fig. 1.2.2 for an enlargement of this region.

This forms the current conducting path of the field effect transistor device based on the layer system. The 'spacer' layer, as its name implies, is a buffer between the channel electrons and the ionized donors. The latter's electrostatic field will scatter the electrons in the channel, reducing their mobility. However, the spacer cannot be made too wide, as this will reduce the channel electron density, which must be maintained at a high level for reasonable device conductivity at room temperature. The final layer shown is that of the 'substrate' or platform on which the growth of the active layers proceeds.

The other layers will also, of course, make their contribution to this model but will be ignored until later for simplicity. The doping mentioned above is found in a planar deposition next to the spacer layer and its presence in the crystal lattice promotes the trapping of a large density of donor electrons. The pinning of the Fermi level on this large trapped density becomes more apparent in the subsequent figures in this section. These trapping sites are called '*DX*' or deep-donor centres. They offer a useful 'boundary' energy in an electrostatic model to be introduced in section 2.1, and also contribute an important effect to electron transport in the channel. Therefore, they warrant an explanatory section of their own - section 1.4.

The substrate is composed of the binary compound gallium arsenide (GaAs), which has become a standard material for supporting the subsequent growth of epitaxial layers. The channel is composed of indium gallium arsenide ($In_yGa_{1-y}As$), which is a ternary material whose relative proportions (y and 1-y) of its group III constituents (In and Ga respectively) can be varied. Variation is crucial to the development of an alloy with a desired band-gap. By substituting indium atoms for gallium atoms, the band-gap (E_G) of GaAs can be decreased approximately

linearly from its value of 1.425eV to the value for InAs (0.36eV). The following equation represents this relationship (room temperature):[12]

$$E_G^{\Gamma} = (1.064(1-y) + 0.36)\text{eV} \qquad (1.2.1)$$

where E_G^{Γ} is the energy band-gap between the valence band maximum and the gamma (Γ) conduction band minimum and *y* is the proportion of indium atoms, which have replaced the gallium atoms. Therefore, if y=0.3, then E_G is 1.105eV. The energy transition across the band-gap is direct for all *y*, that is, there is no accompanying change in momentum. The proportion of In to Ga in this work was either 20% or 30%, therefore, *y*=0.2 or 0.3 respectively.

The spacer layer is composed of the ternary - aluminium gallium arsenide ($Al_xGa_{1-x}As$), where, on this occasion, the substitution of Ga atoms was done using Al atoms. (The energetics of the distortion of the crystal lattice resulting from this substitution will be discussed as part of section 1.4 on *DX* centres) The proportion *x* of Al to Ga is commonly 30% for material to be used as a spacer. This compound has the further benefit that the energy of electrons trapped in deep-donors sited in this material, is known to be around 70meV below the conduction band-edge. The relationship between the band-gap energy and the proportion (x) of the Al to Ga atoms is given by:[12]

$$E_G^{\Gamma} = (1.247x + 1.425)\text{eV} \qquad x \le 0.45 \qquad (1.2.2)$$

Therefore, if *x*=0.3, then E_G is 1.799eV. The band-gap is indirect for x>0.45 where other non-linear relations between E_G and *x* exist.

The different band-gaps (E_G) engineered for these layers can be seen in fig. 1.2.1. When compared side-by-side, the differing band-gaps seem to induce energy offsets in the conduction (E_C) and valence (E_V) band-edges. However, this is not the whole story, as the local effect on the band-edges of these offsets cannot be determined without invoking

two other crucial parameters. These parameters are the vacuum level and the electron affinity. The electron affinity ($e\chi$) of a material, where e is the electronic charge and χ is the potential difference, defines the energy required to overcome the binding energy of a conduction band-edge electron and consign it to 'freedom' with zero kinetic energy. This higher energy level is called the vacuum level. It is very similar to the work function - only differing in its lower reference energy, which for the latter is the Fermi level rather than the conduction band-edge. Some texts denote the symbol χ as the energy[12] and some,[13] as in this work, the voltage.

As can be seen in fig. 1.2.1, this energy is material dependent and can be relatively large. The vacuum level is used as a reference energy for the system. With that reference included, the true relative magnitudes of the conduction (ΔE_C) and valence (ΔE_V) band-edge energy offsets can now be seen. Fig. 1.2.2 shows an enlarged view of these offsets, and clearly the conduction band-edge energy offset (ΔE_C) can be calculated from the following expression:

$$\Delta E_C = e(\chi_1 - \chi_2) \qquad (1.2.3)$$

where $e\chi_j$ represents the electron affinity of the j^{th} material.

The similar offset for the valence band-edges (ΔE_V) additionally depends on the relative energy band-gaps.

If the materials are now 'joined' and a layer contains some net charge, then at equilibrium, the Fermi level should equate in energy across the junctions. The band-edges will vary in order to accommodate this equivalence. If doping electrons exist in the $Al_{0.3}Ga_{0.7}As$ layer at its edge, as stated earlier, then a sizeable density will be trapped in the deep-donor centres approximately 70meV below the conduction band-

edge. The Fermi level, therefore, is pinned on this energy at this position. As a degenerate electron gas exists in the channel layer of such systems, it can be further assumed (and confirmed) that the Fermi level

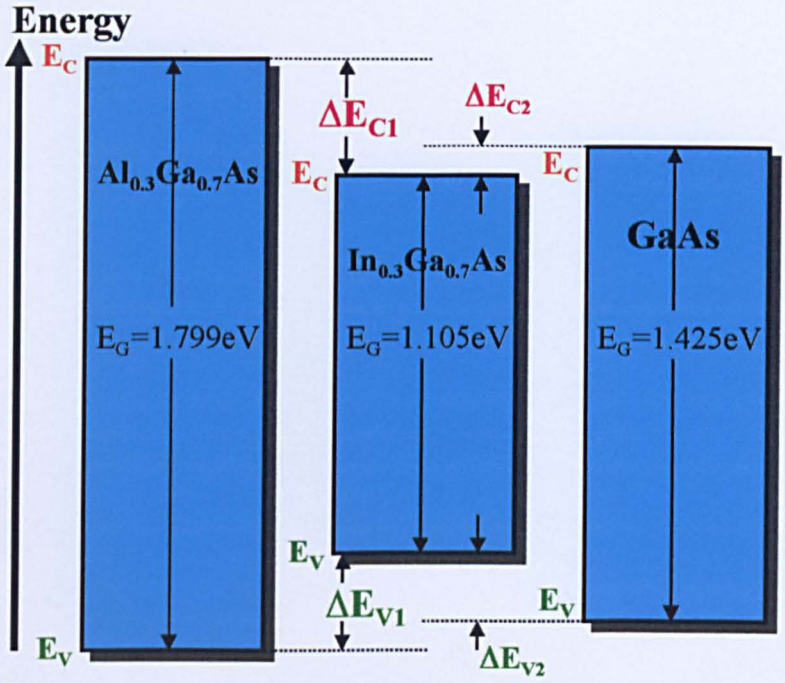


Figure 1.2.2 An enlarged section of part of figure 1.2.1 around the band-gap regions of the three isolated semiconductors. Note in particular the energy offsets, ΔE_{C1} and ΔE_{C2} that will, eventually, define the electron channel borders in energy terms. Their values are 429meV and 110meV respectively. (calculated from values given in ref. 12)

will lie at some energy above the conduction band-edge over this locality. It would also be expected (and confirmed) that ideally, the Fermi level would extend eventually to a mid band-gap energy deep into the undoped substrate. Fig. 1.2.3 is the result of such expectations and assumptions.

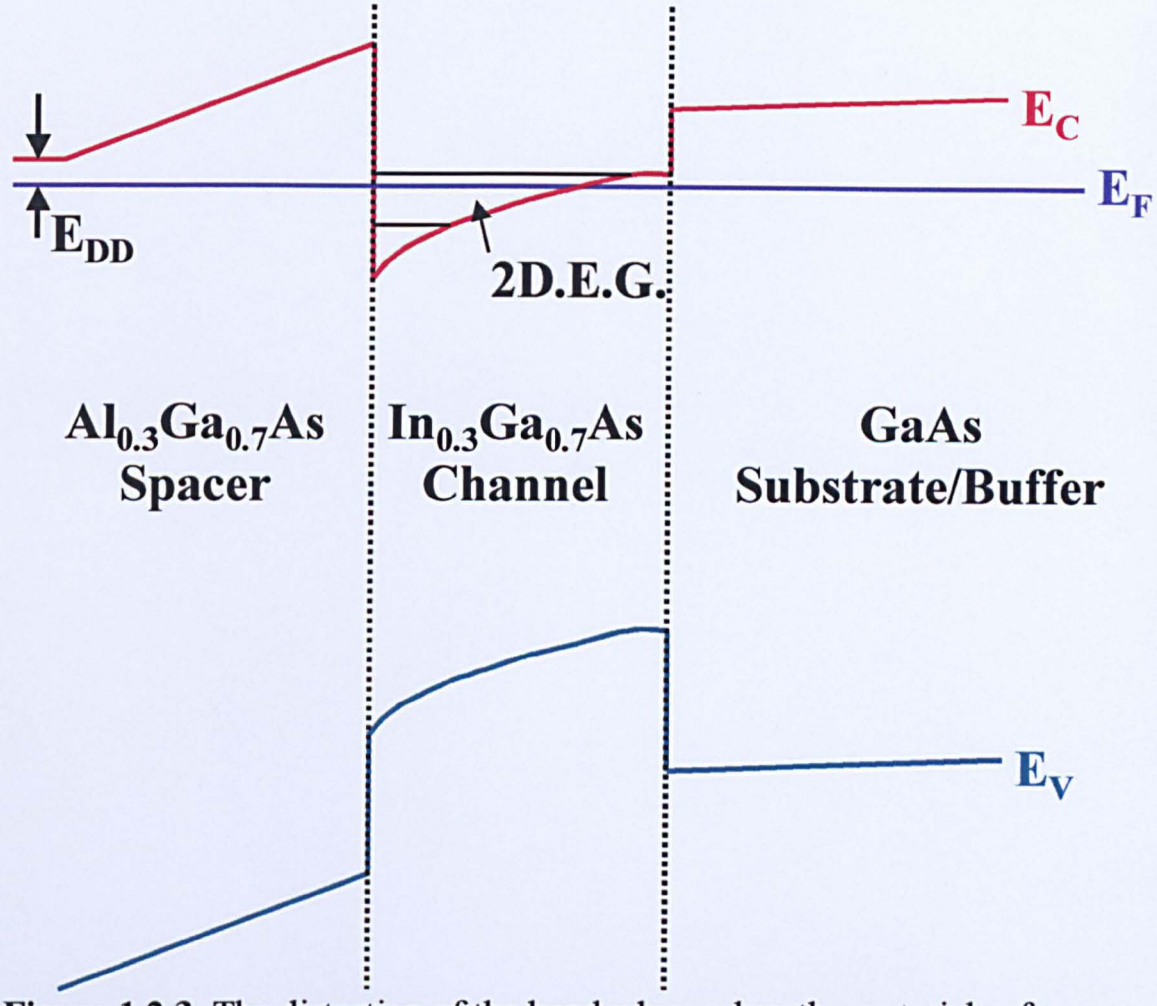


Figure 1.2.3 The distortion of the band edges when the materials of figures 1.2.1 and 1.2.2 are 'joined'. Note the continuation of the Fermi level (E_F) across the layers, with its pinning at energy E_{DD} (energy of deep-donors) below the conduction band edge (E_C). As will be seen in section 1.4 and figure 1.2.4 in section 1.2, this pinning of the Fermi level relative to the conduction band edge is due to a large trapped population of electrons in the deep donors found there. The relative energy of the Fermi level elsewhere requires an additional self-consistent calculation because the spatial distribution of charge requires knowledge of the band profile, which in turn demands knowledge of the charge distribution. Note also that the Fermi level lies in the first subband in the channel and hence the electrons trapped here are degenerate - the 2 dimensional electron gas (2DEG).

1.2.3 The Limitations of the Energetics Perspective

Clearly, the less than certain language evolving in the last sub-section indicates that this simple visual model is limited in its capabilities. To determine the relative energy difference between the Fermi level and the conduction band-edge across all the layers requires a self-consistent calculation. Such a calculation solves, simultaneously, equations that express the electrostatics and the quantum mechanics of the situation. The simultaneity is required as the spatial charge distribution depends on the energetics - expressed through the band-edges. However, the energetics, in turn, depend on the charge distribution. Therefore, there is a mutual interdependence on the conditions and solutions of both the Schrödinger and Poisson equations. More on the implications of this result will be given throughout this chapter. By making some reasonable approximations however, some useful parameters can be determined by applying only electrostatic principles. The resulting model is presented in section 2.1.

In order to accommodate the constant Fermi level, the band-edges have to vary across the layers, as stated earlier. The energy variations are linear with depth in sections where no doping or charge exists. Over these sections the electrostatic field is constant - created by opposing charges of equal magnitude situated at either end of the section. Where there is a distribution of charge with depth, for example in the channel, the field is no longer constant and hence, the band-edge is non-linear there. More detailed and readable descriptions of the electrostatics across heterojunctions can be found in Long et al[14] (1993), Skuras et al[15] (1995) and the text book by Davies[16] (1998). Of course, extensive reference will be made to these works in later sections. A numerical

version of the self-consistent calculation is available as a Macintosh/PC program, which is freely distributed by Prof. Greg L. Snider, University of Notre Dame.

1.2.4 The Full Conduction Band-Edge Energy Profile

The full energy profile of the conduction band-edge is shown in fig. 1.2.4. For this diagram, the channel is within a 'deep' system,

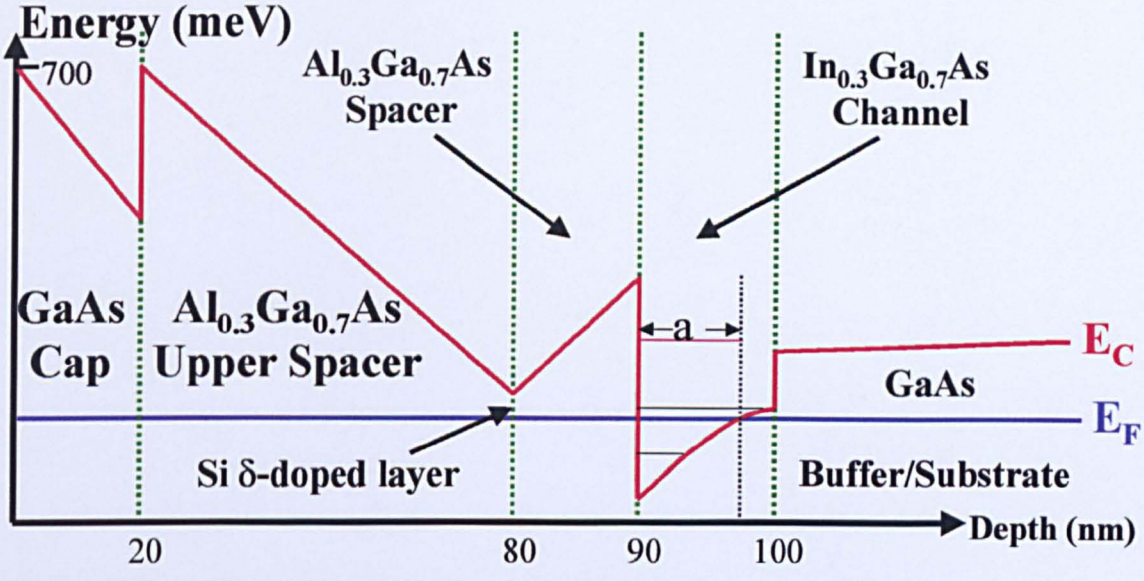


Figure 1.2.4 The full conduction band-edge energy profile, now extended from fig. 1.2.3 to include the upper layers. The example shown has a 10nm spacer, with 30% indium concentration in the channel, which is 'deep'. The relative energies of the Fermi level and the conduction band-edge require self-consistent calculations, but to a good approximation it may be taken as pinned at 70meV below the conduction band in the Si δ-doped layer. Ordinarily, there is a significant population of electrons trapped in the *DX* or deep donor centres there. The GaAs monolayers, around the δ-layer are not shown. An effective channel width (a) is introduced and an explanation is given in the text below.

with its indium content set at 30%. A 'deep' system is that denoted by the thicker cap and upper spacer values as shown in table 1.1.1. The spacer thickness is set to 10nm. This profile is often seen in the

literature,[17] but it was confirmed using the Snider program when that became available. Note how the addition of the other layers has affected the energy distributions of the conduction band-edge and the Fermi level. One interesting feature is that the Fermi level does not lie above the conduction band-edge right across the channel. We therefore use the concept of an effective channel width (a), see fig. 1.2.4. It is difficult to determine exactly this effective width, as it will vary slightly with the parameters of the layer system. This will be so despite the physical width of the channel remaining constant. However, a value for a of (7±1)nm is reasonable for a physical channel width of 10nm as here.

In order to model the electrostatics of these systems, it was advantageous, but not necessary, to pin the Fermi level on the electrons trapped in the *DX* centres. This was encouraged by the incorporation of a thick buffer layer below the channel to make its pinning on the charge in the neighbouring superlattice less probable. Most electrons originating from the donor layer get trapped in the surface states of the GaAs cap. The energy of these states is mid band-gap, that is, about 700meV below the conduction band. The addition of a metal gate provides a Schottky energy barrier of about the same value, therefore, the Fermi level can be regarded as pinned on this energy.

The channel and spacer width never exceeded 10nm. The former to eliminate possible strain relaxation and the latter because the channel density would become too low for realistic device use. The upper spacer merely provides a suitable semi-insulating separation between the gate and the active layers below. In devices, this layer is made as narrow as possible to avoid fringing effects at the channel due to the gate field, and to lower threshold gate voltages. However, in D.C. measurements, gate leakage often arose with the use of such a narrow layer, and therefore

for most studies, a thick upper spacer layer or 'deep' channel was grown.

The GaAs cap was grown for two main reasons:

1 The adhesion of the gate (or ohmic) metal to GaAs is better than it is to AlGaAs.

2 The energy band-offset between the cap and the upper spacer prevents the formation of holes by the migration of electrons from the cap layer due to a subsequently applied repulsive gate field.

The energy offset (ΔE_C) between the spacer and the channel is around 430meV when a 30%, and 360meV when a 20% indium content exists in the channel. For the former concentration, the offset between the channel and the GaAs buffer is 110meV.

A self-consistent approach demands that all the layers and their associated parameters impart a greater or lesser contribution to the resulting spatial charge density distribution. Intuitively however, one would expect that only the layers shown in figs. 1.2.1 to 1.2.3 and their parameters, along with the doping density, would primarily influence the carrier density of the channel. It turns out that this is also the case, when simple electrostatics is applied.[14 - 16] Surprisingly the doping density can be disregarded in the first order provided it is high enough because, as stated earlier, the vast majority of its electrons populate the surface states of the (100) GaAs cap and because the high density of states in the delta layer leads to a weakly varying Fermi energy.

1.3 The Accommodation of Lattice Mismatch by Pseudomorphic Growth

1.3.1 Introduction

The implication of the last section on heterojunction formation is one of complete freedom in band-gap engineering - a ternary compound could be used with any value of its energy band-gap provided it lay between the band-gaps of its binary end-points. For example, the ternary compound $Al_xGa_{1-x}As$ could have any value between that of GaAs (1.425eV) and that of AlAs (2.1eV), although there is a transformation from a 'direct' to an 'indirect' bandgap at about x=0.45. Likewise, the band-gap of $In_yGa_{1-y}As$ could lie between that of GaAs and that of InAs (0.36eV). The heterojunction formed between such ternary compounds could, therefore, generate a wide range of conduction band-edge energy offsets (ΔE_C).

These statements are true in principle, but they take no account of the possible problems arising from the relative atomic mismatch between the materials either side of a junction. As will be seen in section 1.4, other factors also conspire to limit the practical exploitation of the full band-gap ranges theoretically available. However, in this section, the limitations enforced by lattice matching will be discussed.

1.3.2 The Equilibrium Lattice Parameter

In elemental Group IV semiconductors, such as silicon (Si) and germanium (Ge), the atoms bond by sharing equally their four valence electrons. In this manner, each atom will have effectively, on average, eight valence electrons and a stable closed shell. Covalent bonding is the result and confers semiconductors with much of their hardness but brittleness. This bonding requires a full quantum mechanical explanation, unlike ionic or even inter-molecular (van der Waals) bonding, where classical electrostatics can allow reasonable approximations. What can be stated, however, is that an equilibrium inter-atomic separation (a_0) will form, on average, between the participating atoms. Qualitatively, this value denotes the separation at which repulsive and attractive forces cancel.

In a III-V compound such as GaAs, there is a slightly unequal sharing of the electrons. The As atom has five and the Ga atom three valence electrons, therefore more electrons will be 'found', on average, nearer the As atom. Hence this bond, although still covalent, will be characterised by a net charge distribution - the end with the As atom has more negative charge, the Ga end will have a relative net positive charge. In contrast with an elemental semiconductor, the equilibrium inter-atomic separation will have an input from the different 'sizes' of both atoms and also this slightly polar-covalent bonding between them.

If now another group III atom, such as indium or aluminium substitutes on a proportion of the Ga sites, then the different 'sizes' of these atoms will also have an input on the eventual value for a_0. No one covalent bond will form in isolation from all the others, which comprise

the crystal. The overall wavefunction will superpose over the whole structure and a_0 will evolve to minimise the potential energy. Clearly then, the proportions and distribution of the substituted element will influence this average atomic separation. Substituting atoms of various sizes in various proportions into the basic periodicity of the growing single crystal network will induce compressive or stretching forces on the local atomic bonding, which will shift commensurately the value for a_0. A characteristic length representative of the translational symmetry of this ternary compound - the lattice parameter (a)- will also be affected.

1.3.3 The Crystallisation of Compound Semiconductors

GaAs crystallises in the 'zincblende' configuration, which is very similar to the 'diamond' formation of the elemental semiconductors. A simple construction,[18] which aids the visualisation of this configuration, is described below. The constituent components, Ga and As are each assigned a face-centred cubic (F.C.C.) structure and these two cubes or sub-lattices are inter-penetrated to a relative position such that a minimum average bonding potential energy at equilibrium exists over the modified structure. The group III Ga sub-lattice could also contain a proportion of other group III atoms such as In or Al, which will slightly alter the lattice parameter of this sub-lattice. At about room temperature, the lattice parameter of GaAs settles at a value of 5.6532Å.[12] If all the Ga atoms are substituted with Al atoms, this value settles to 5.6611Å - a difference of only about 0.14%. Although the Al atoms strain the local atomic environment, the overall effect of its

substitution causes hardly any change in the lattice parameter. If however, indium atoms substitute, a significant change in the lattice parameter evolves. Complete substitution culminates in InAs with a lattice parameter of 6.0584Å at about room temperature. Compare this value with that of GaAs and a significant percentage change of 7.17% would have to be accommodated.

1.3.4 Lattice Mismatch-Induced Strain.

In isolation these compounds can distribute any strain over the whole crystal bulk, and it is thus seldom a problem. However, if an abrupt heterojunction exists between materials whose lattice parameters differ significantly, there can then ensue strain-induced problems. If $Al_xGa_{1-x}As$ is grown onto GaAs then, as stated above, the maximum mismatch is only 0.14%. This is unlikely to induce any damaging strain across the junction, and experimentation bears this out.

In order to achieve, and hence exploit, high channel densities, a large conduction band-edge energy offset is required between the channel and the spacer. An $Al_xGa_{1-x}As$ spacer grown abruptly onto an $In_yGa_{1-y}As$ channel layer achieves this, see section 1.2. However, at larger values of y, the relative lattice mismatch at the junction may become excessive. (The same principle also applies to the substrate-channel junction.) A further practical limitation exists in that only substrates composed of GaAs or indium phosphide (InP) are generally available with a low defect density. Any projected exploitation of a layer material should account for this restraint.

The GaAs substrate/buffer used in this work is grown with the [100] plane exposed and, as such, would have the appearance of fig.1.3.1[19] from above. The Ga and As atoms lie in different planes and each atom in the upper plane forms two bonds with two different atoms in the lower plane. This will leave, on average, two 'dangling' bonds per atom at the surface to which new epitaxial growth will bond. This crystalline structure will, of course, be extended to all subsequently grown layers. The only difference being that other group III atoms will replace some Ga, accompanied by the commensurate change in lattice parameter.

Between the values of the two end-point binary compounds, the lattice parameter varies linearly with the mole fraction. This relationship for $In_yGa_{1-y}As$ is expressed through Vegard's Law:[12]

$$a_{(In_yGa_{1-y}As)} = ya_{(InAs)} + (1-y)a_{(GaAs)} \qquad (1.3.1)$$

where $a_{(j)}$ is the lattice parameter of the undernoted compound (j). Therefore, if y=0.3, $a_{(InGaAs)}$ = 5.77Å, and the lattice mismatch with GaAs is about 2%. The $In_{0.3}Ga_{0.7}As$ channel, when grown on the GaAs substrate, will align with the schematic configuration shown in fig.1.3.1, but will be characterised by an average dimension, which is 2% larger. This may seem small, but it is applied abruptly at the heterojunction, and thus, the local chemical bonds will be significantly strained. If relaxation of this strain develops, the band-edge offset remains but performance-degrading dislocations will disperse throughout the channel.

The percentage mismatch values quoted previously were obtained from the following simple formula:

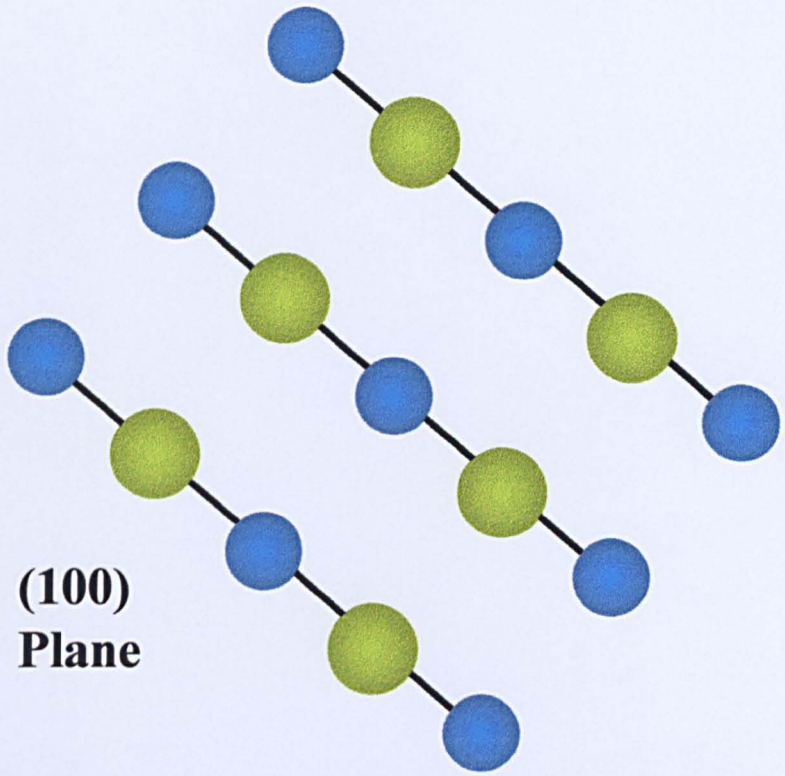


Figure 1.3.1 The plan view of the growing GaAs substrate, where the [100] direction of growth is chosen. Either coloured sphere can represent Ga or As, but one layer (colour) lies in a plane below the other. The growing plane of a subsequently grown ternary will look much like this with some Ga atoms replaced with In or Al atoms.

$$\frac{\Delta a}{a} = \frac{100(a_T - a_B)}{a_T} \qquad (1.3.2)$$

where a_T and a_B are the lattice parameters of the top and bottom layer respectively. In the case of an $In_{0.3}Ga_{0.7}As$ (top layer) grown onto GaAs (bottom layer) this ratio turns out to be positive as $a_T > a_B$. This implies that the bonding in the $In_{0.3}Ga_{0.7}As$, particularly close to the junction, will be under biaxial compressive strain, whereas the GaAs will be affected by biaxial tensile strain. However, if one layer is much thicker than the other, for example the GaAs substrate, then it is the atoms of

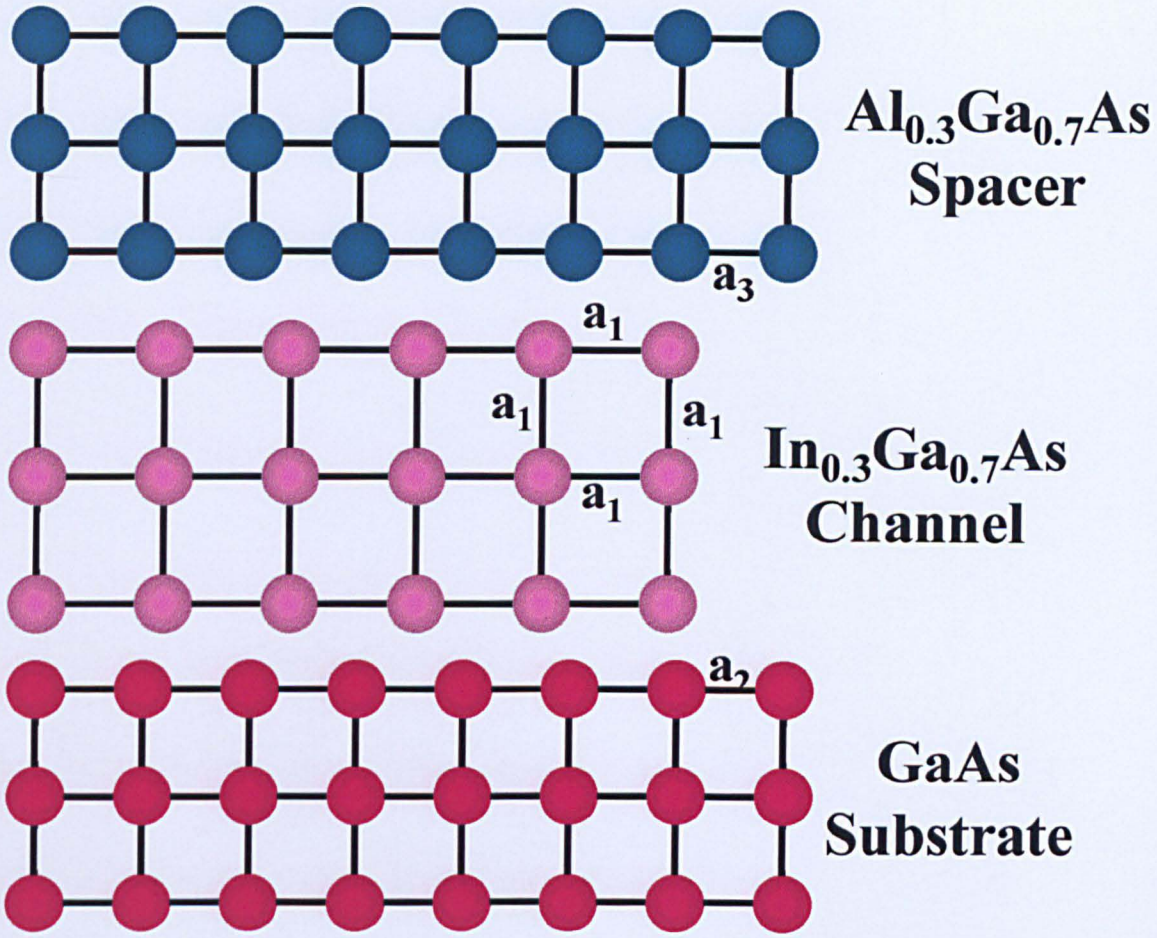


Figure 1.3.2a A schematic representation of the substrate, channel and spacer- each in isolation showing the lattice mis-match. The values of a_2 and a_3 are very similar, with a_1 different. The relative differences in lattice parameter are much exaggerated for visibility. The layer widths shown are, of course, much thinner than in reality.

the thinner layer, which will undergo most of the displacement and hence strain. It is the lattice parameter ($a_{\|}$) parallel to the growing surface, which is reduced in the case of $In_{0.3}Ga_{0.7}As$. To allow for this, the perpendicular lattice parameter ($a_{\perp}$) expands by what is known as the Poisson effect. This is schematically shown in figs. 1.3.2a and b, where the subsequent addition of the $Al_{0.3}Ga_{0.7}As$ spacer layer, with its smaller lattice parameter, is also shown.

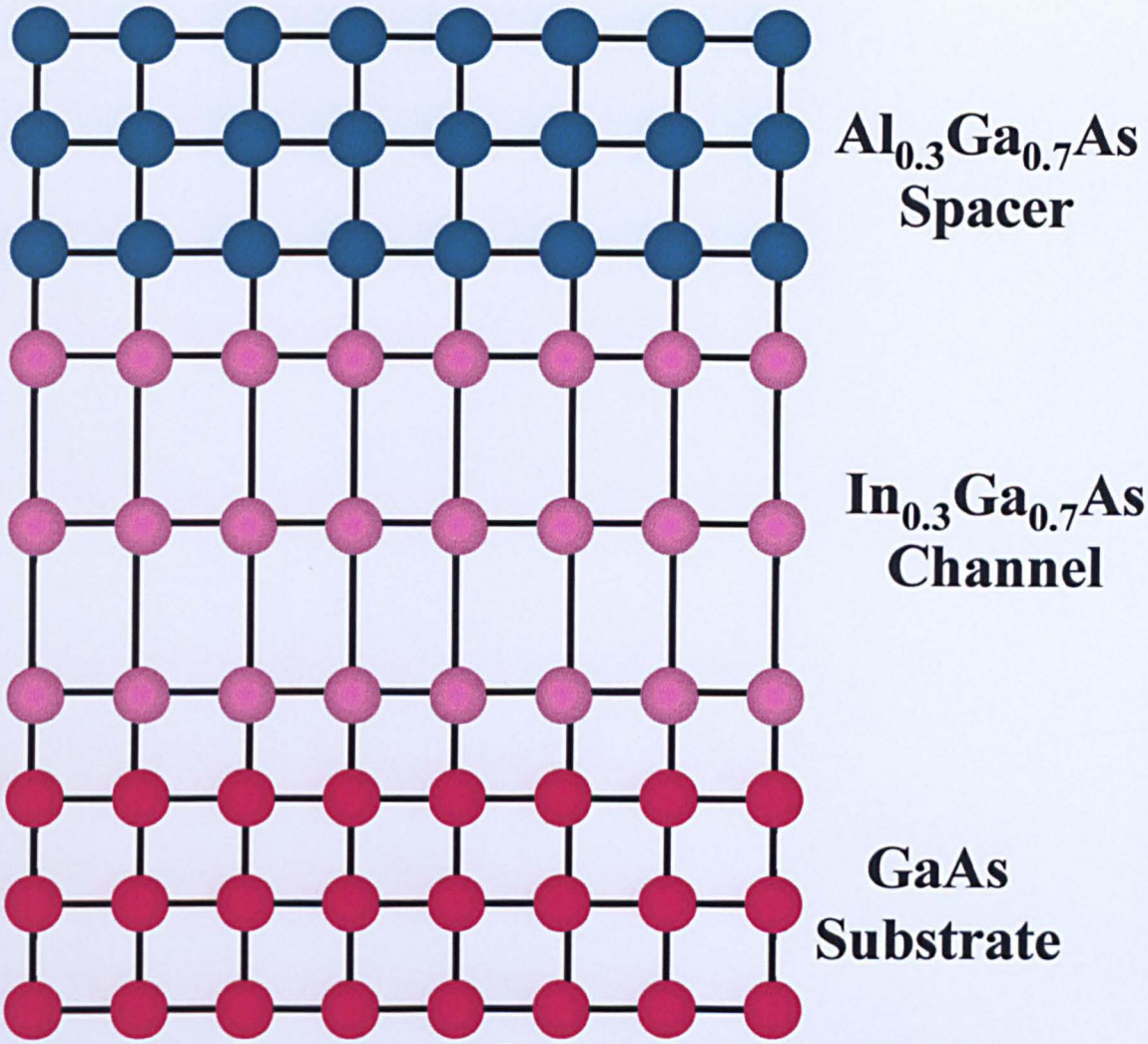


Figure 1.3.2b The extension to fig. 1.3.2a showing the effect of the pseudomorphic strain on the channel layer in particular. As the substrate is assumed infinitely thick then most of the change will occur in the lattice parameter of the channel, that is, a_1 tends to a_2. To compensate there is a tensile strain applied to the bonds in the vertical direction (Poisson effect). These strained vertical bonds may also influence the bonding within the spacer layer above.

the subsequent addition of the $Al_{0.3}Ga_{0.7}As$ spacer layer, with its smaller lattice parameter, is also shown.

Depending on the actual width of the $In_{0.3}Ga_{0.7}As$ channel layer, it would seem that $a_{||}$ could persist through to the spacer, which would enable the $Al_{0.3}Ga_{0.7}As$ to lattice match without much strain. The perpendicular tensile strain on the $In_{0.3}Ga_{0.7}As$ side may however have some effect. It seems likely therefore, that most of the strain will be

accommodated or generated at the substrate-channel boundary, where the growing layer ($In_{0.3}Ga_{0.7}As$) adopts approximately the value $a_{||}$ of the GaAs substrate below. This type of growth is called 'pseudomorphic', because the morphology of the lower layer is adopted. The parallel lattice parameter ($a_{||}$) for the InGaAs can be calculated from the following equation:

$$a_{\parallel} = \frac{a_{InGaAs} d_{InGaAs} + a_{GaAs} d_{GaAs}}{d_{InGaAs} + d_{GaAs}} \tag{1.3.3}$$

where a_j and d_j is the unstrained parallel lattice parameter and the thickness respectively of the undernoted layer (j). Strictly, this equation and those following only apply to growth along the [100] direction. Note that as $d_{GaAs} >> d_{InGaAs}$, then $a_{||}$ will tend to a_{GaAs}, as implied above. Re-arranging equation 1.3.2 and allowing for the above limit gives:

$$a_{\parallel} = \left[1 - \left(\frac{\Delta a}{a}\right)_{InGaAs}\right] a_{InGaAs} \tag{1.3.4}$$

The perpendicular lattice parameter ($a_{\perp}$) of the InGaAs is given by:

$$a_{\perp} = \left[1 + \frac{c_{11}}{2c_{12}}\left(\frac{\Delta a}{a}\right)_{InGaAs}\right] a_{InGaAs} \tag{1.3.5}$$

where c_{11} and c_{12} are elastic constants of the InGaAs layer.

A 3x3 matrix can represent the full 3-D nature of the stress (S) in the following equation, which is a form of Hooke's Law:

$$\begin{bmatrix} S_{xx} \\ S_{yy} \\ S_{zz} \end{bmatrix} = \begin{bmatrix} c_{11} & c_{12} & c_{12} \\ c_{12} & c_{11} & c_{12} \\ c_{12} & c_{12} & c_{11} \end{bmatrix} \begin{bmatrix} f_{xx} \\ f_{yy} \\ f_{zz} \end{bmatrix} \tag{1.3.6}$$

where f now substitutes for $\Delta a/a$. The x-y plane is the growth plane and the z direction is the direction of growth.

As $f_{xx}=f_{yy}$ (the two orthogonal mismatches in the growing plane are equal), both defined as $f_{||}$, and the stress in the z direction is zero then:

$$f_{zz} = \frac{-2c_{12}}{c_{11}} f_{||} \qquad (1.3.7)$$

For the case of pseudomorphic growth, the strain energy per unit area (E_{ST}) is given by:[20]

$$E_{ST} = \left[\frac{1}{2}c_{11}\alpha + c_{12}\beta + \frac{1}{2}c_{44}\gamma\right]d \qquad (1.3.8)$$

where $\alpha=f_{xx}^2+f_{yy}^2+f_{zz}^2$; $\beta=f_{xx}f_{yy}+f_{yy}f_{zz}+f_{zz}f_{xx}$ and $\gamma=f_{xy}^2+f_{yz}^2+f_{zx}^2$. d is the thickness of the InGaAs layer here. The equation (1.3.8) also assumes that the substrate is infinitely thick. For crystal orientations where the normal to the growing surface is also an axis of symmetry, then $f_{ij}=0$ where $i\neq j$. Using equation 1.3.7 and the equality expressed before that equation, then equation 1.3.8 simplifies to:

$$E_{ST} = f_{||}^2\left(c_{11} + c_{12} - \frac{2c_{12}^2}{c_{11}}\right)d \qquad (1.3.9)$$

That is, $E_{ST} \propto d$, and $f_{||}^2$ (or$\Delta a_{||}^2$). This relation is critical for the successful growth of pseudomorphic layers and much research and experimentation has been conducted to determine a critical thickness (h_c) of d whereupon the strain relaxes.[21] If h_c is not exceeded, the minimum state energy can be accommodated by strain. Above h_c, however, this state can be attained only by the formation of dislocations, where the bonding re-arranges amongst the atoms to reduce the potential energy. Fig.1.3.3 displays the atomic environment around the formation of a dislocation. Once dislocations appear at the mismatched interface, they can spread upwards with the growing layer. As they propagate or 'thread' they can multiply and achieve a very high areal density of $>10^9 cm^{-2}$. A theoretical prediction for h_c has been given as:[12]

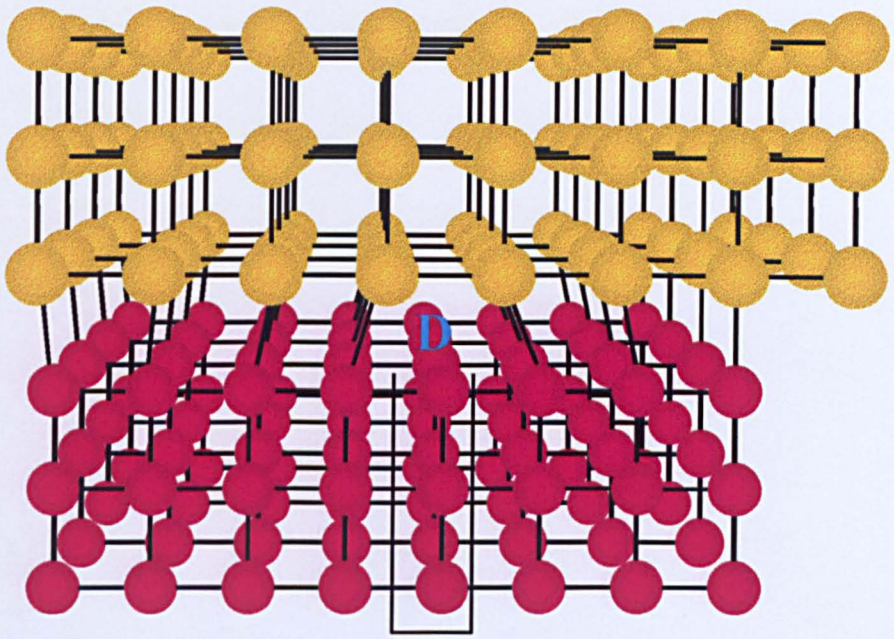


Figure 1.3.3 The formation of a dislocation (D), and the extra plane of atoms (shown bordered). By this re-arrangement of atoms, the lattice can reduce the potential energy stored in its bonds. These dislocations can however severely degrade device performance.

$$h_c = \frac{a_{GaAs}\left(1-\frac{\nu_{PR}}{4}\right)\left[\ln\left(\frac{h_c\sqrt{2}}{a_{GaAs}}\right)+1\right]}{2\sqrt{2}\pi|f|(1+\nu_{PR})} \qquad (1.3.10)$$

where ν_{PR} is the Poisson Ratio ($c_{12}/(c_{11}+c_{12})$) and is about 1/3 for most semiconductors and f is $\Delta a/a$. A GaAs substrate or host layer has been assumed. Note that h_c appears on both sides of this equation, therefore an iterative approach to its solution is required. Considering energy minimisation, a simpler approximation for h_c is possible:[12]

$$h_c \approx \frac{a_{InGaAs}}{2f} \qquad (1.3.11)$$

A great deal of variation for h_c exists in the literature, and it is known now that the growing condition in general and lower growth temperatures in particular, can extend h_c somewhat. A good 'rule of

thumb' is that quality growth of pseudomorphic layers is possible if the product of the layer's thickness in Å and its percentage mismatch is less than 200Å%.[3]

Lattice mismatching with its accompanying pseudomorphic growth, offers band-gap engineering an extended 'window'. The presence of strain however, does alter, to a greater or lesser extent, some of the crucial parameters of the affected material(s) from their values determined in the absence of strain. The band-gap and the band structure can be slightly modified under this magnitude of strain. Variation of the effective mass of the charge carriers with indium content and hence strain will also occur.[22] In addition, how does the presence of strain affect the actual distribution of atoms under epitaxial growth? Will its presence promote additional disorder, or perhaps even additional clustering of a given atomic species? These questions will arise again when molecular beam epitaxial growth and electron scattering are discussed later.

1.4 Delta (δ)- Doping and *DX* Centres

1.4.1 Introduction

As the existence of the so-called *DX* or deep-donor centres is important to much of the analysis of the experimental data, we now devote some time to describe them. As an example, the derivations of the electrostatic model,[14,15] which is described in section 2.1, depend on the Fermi level pinning on a large electron density trapped in such centres. In general, the density and mobility of electrons in the channel are influenced by the electronic occupation or otherwise of the remote *DX* centres. The creation of *DX* centers arises as a side effect of the doping procedure, and therefore this section will commence with a brief discussion of delta-doping, which is the procedure of choice in this work.

1.4.2 Delta- (δ)-Doping

In order to supply the channel with sufficient electrons, a doping layer must be incorporated during the epitaxial growth stage. For the layer systems considered here, the doping atoms are of group IV silicon (Si) and they are deposited in a single plane – a technique known as delta-(δ) doping. δ-doping confers a number of advantages over slab-doping, which distributes the dopant atoms and hence, the donor electrons over a sizeable volume of the material. They include:

1 Their narrower distribution with layer depth promotes easier modelling of a number of characteristics. For example, the energy variation of the conduction and valence band-edges, perpetrating through the layer system, will largely be linear (fig. 1.2.4 in section 1.2.4).

2 All the donor electrons have, in principle, the same opportunity of populating the channel. Higher charge density, and hence conductivity should be possible, which is desirable for device purposes. Indeed, Ensemble Monte-Carlo simulation[23] predicts that δ-doping not only increases channel carrier density but also electron drift velocities.

The closer the δ-doping layer is sited to the channel, the greater the likelihood of a larger charge density there. Much research has been done[24] in optimising the relative positions of the channel and the doping layer, including some studies of this type in the present work. However, with a large population in the channel, higher subbands could be occupied, which will generally reduce the overall mobility. Also, scattering of the channel electrons by the donor ions may become excessive, which is likely to counteract the advantage of higher carrier density.

Although a single areal density of the Si dopants is the ideal, in reality there will be some spreading in directions perpendicular to the layers. Many methods for limiting, or at least controlling, the magnitude and direction of the spreading tendency have been studied. The temperature of the substrate controls much of the spreading. Substrate temperatures are usually high (>500°C) and this factor will inevitably encourage an enhanced spreading. The effect of varying substrate temperature, and hence spreading rates, before and after doping, have

been studied. The purpose, of course, is to determine an eventual growth sequence, which will limit, or at least render predictable, this spreading or surface segregation.

1.4.3 The Formation of *DX* Centres

Doping a binary compound such as GaAs with a group IV species such as Si is a relatively simple process to model. There is a straight substitution of the Ga atom with a Si atom, with some elastic strain induced in the covalent chemical bonding, and the donor electrons would normally lie in hydrogenic 'shallow' states.

Compare that relative simplicity with the situation that can evolve with similar doping of a ternary compound such as $Al_xGa_{1-x}As$. This host lattice already accommodates some strain due to the Al substitution in (100 times x)% of the former Ga sites. To obtain donor electrons, the additional incorporation of Si into this complicated lattice is necessary.

When this was done experimentally, initial studies revealed that, in direct comparison with GaAs, free electron densities were lower, but persistent photoconductivity was found. The conductivity arising from the electrons was temperature dependent. Lang and Logan,[25] and Lang et al,[26] in the late 1970s, proposed that, on the capture of an electron, the lattice relaxed in the proximity of a donor. About this time, the term *DX* was concocted for the relaxed defect, with the '*D*' representing a donor, and the '*X*' for some unknown defect - together comprising a defect complex. The *DX* centres were largely perceived as a nuisance and methods were proposed to eliminate their effect. However, the use of $Al_xGa_{1-x}As$ as a spacer layer has many desirable features, and

therefore, the characterising and modelling of these 'complexes' gained momentum.

If hydrostatic pressure is applied to Si doped GaAs, then many of the *DX* characteristics of $Al_xGa_{1-x}As$ appear,[27] including persistent photoconductivity. Infra-red measurements confirmed, through observation of the vibrational modes of Si in GaAs, that the Si was largely incorporated by substitution. Therefore, even GaAs can support the existence of a complex *DX* centre if the pressure is appropriate. This centre was found[28] to lie at an energy of 260meV above the conduction band-edge.

This, and other experimental evidence, strengthened the case that $Al_xGa_{1-x}As$ contains electronic states of occupation in isolated donor atoms, and was supported by the additional observation of hydrogenic shallow states originating from the same species. These are the states associated with an isolated donor atom and not a defect complex, when that donor substitutes for a group III atom.

Experimentation[29 - 32] has revealed that four *DX* centre types or levels exist in Si-doped $Al_xGa_{1-x}As$. Deep Level Transient Spectroscopy (DLTS) has found that multiple signal peaks of varying strength exist in $Al_xGa_{1-x}As$ at different temperatures for a given value of x. As expected only one is observed for GaAs. The strongest peak is independent of x, hence cannot be the result of a transition between the deep levels and a different conduction band minimum. The conclusion emerges that the Si donors are hosted in varying concentrations of different local atomic environments,[30 - 34] each with their own energetics. This is not surprising when all the local sites possible in a complicated ternary compound are considered.

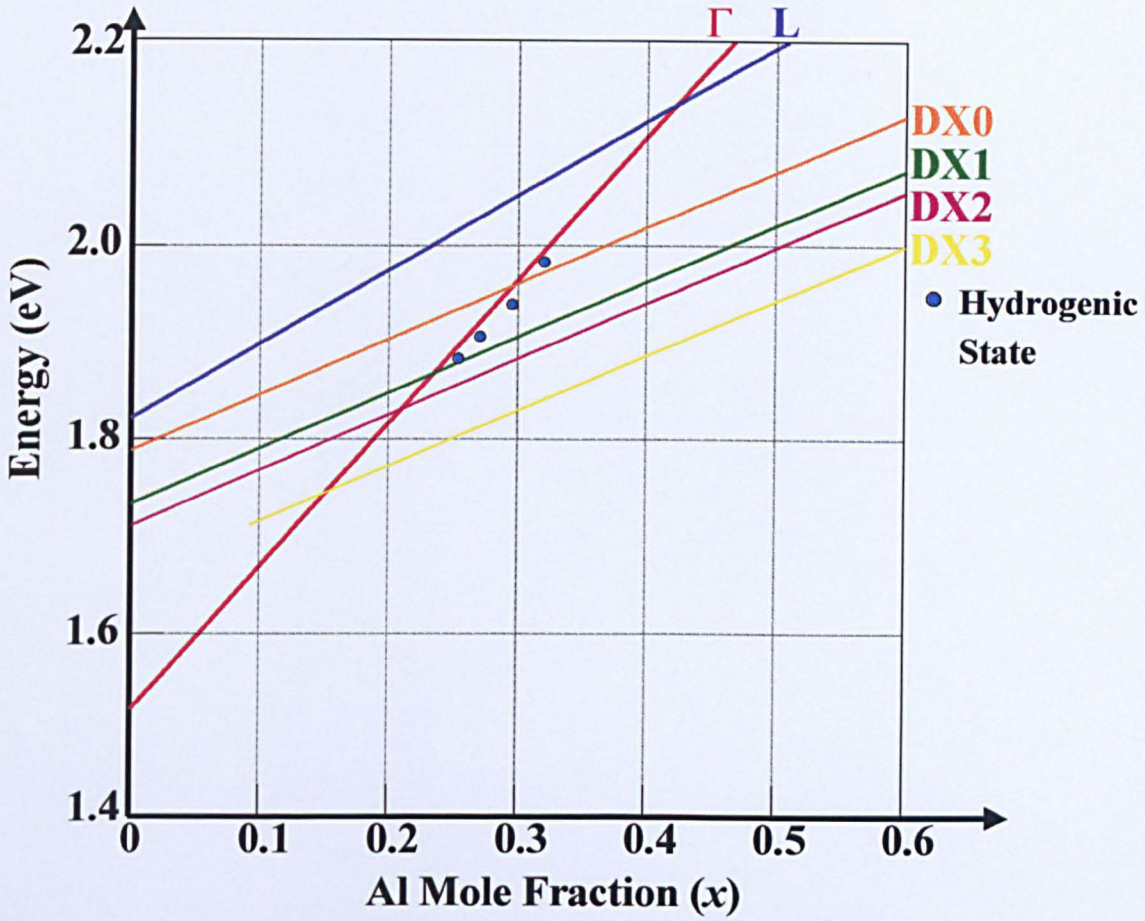


Figure 1.4.1 The energy relationship between the aluminium mole fraction and the energies of the various *DX* centres and the hydrogenic states.[33] Note, in particular, the comparative values at x=0.3. It should also be remarked that the density of the *DX* centres decreases with both increasing assigned number (0,1,2 and 3) and decreasing x[35].The variation with x of the energy of the *DX*0 centre is known,[36] but it is an assumption that the other lines lie parallel, although the energy difference at some x values is known.

The relationship between the Al mole fraction and the energies of the L and Γ conduction band minima and the four known *DX* types is shown in fig. 1.4.1. Also shown for comparison are some hydrogenic levels. Note, in particular the situation at x=0.3, and compare the Γ-minimum with the hydrogenic and the various *DX* types. By substituting atoms other than Al in the GaAs lattice and observing the perturbations, it was concluded that the local atomic configuration must play a role, in

addition to the isolated donor states, culminating in distorted donor configurations.

1.4.4 Modelling the *DX* Centre

Based on much of the evidence related above and much more published elsewhere, Chadi and Chang proposed a model for the *DX* centre which has gained wide acceptance. The key publication is their 1989 paper,[36] which builds on their own previous work in 1988.[37] They have the Si atom moving along the [111] crystal axis after its bond with one of its As neighbours is broken. It finds an interstitial site close to three other group III atoms. In 2-D cross-section, it may look something like fig. 1.4.2.[16] A 3-D version can be seen in refs. [36] and [38].

Clearly, there are four different possibilities for the atomic configuration of group III atoms around the donor - zero, one, two or three Al atoms. If the Al mole fraction is low, then the 3 Al neighbour situation, that is, the *DX*3 centre, should be relatively rare. The caption to fig. 1.4.1 implied this, and low-strength DLTS signals emanating from some *DX* types confirmed this. Statistics suggest that the *DX*2 and *DX*3 centres trap the most electrons, although some dependency on x does occur. Their electron occupation is primarily dependent, for a given temperature, on the relative energy between their ground states and the Fermi level. If doping is sufficiently high and the temperature permits, *DX* occupation can be large enough to pin the Fermi level.

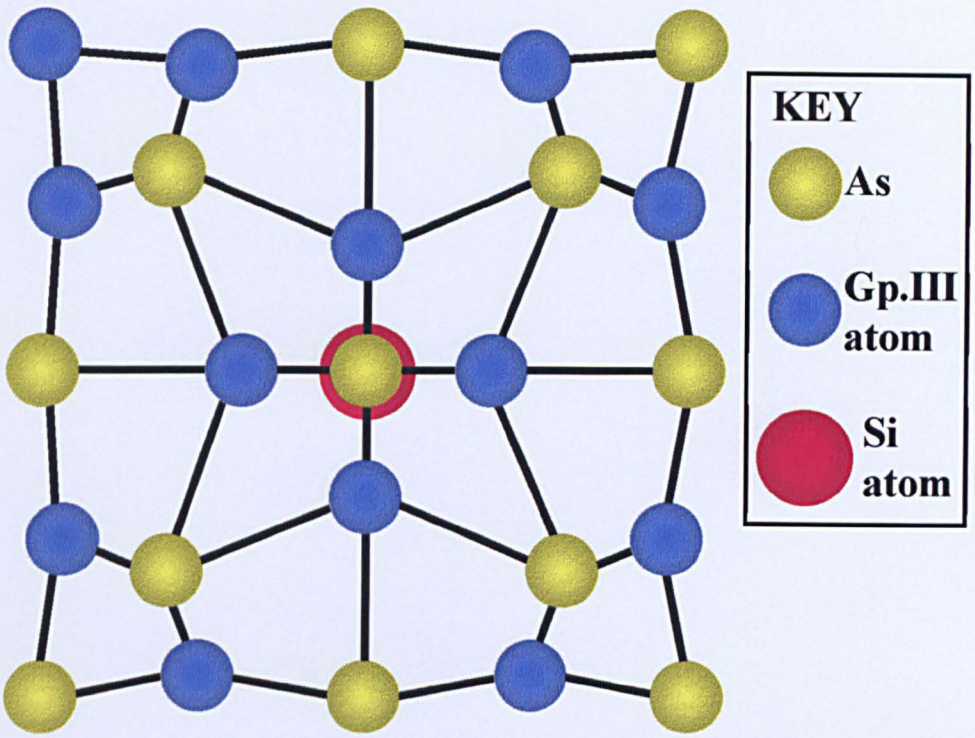


Figure 1.4.2 Schematic 2-D cross-sectional view of a DX centre, showing the lattice distortion. The blue circles are group III atoms, the yellow circles are As atoms and the red circle is a donor atom/ion. (after ref.[16])

One further prediction of Chadi and Chang was that distorted configurations were able to stabilise with the capture of two electrons. This would imply that there are now at least three possible electronic states for a given *DX* centre type. All three contain the positive donor ion, but vary whether they contain zero, one or two electrons. The zero electron condition is known as the D^+ or d^+ state, which has no trapped electrons. A single positively charged entity thus remains. The one electron condition is known as the D^0 or d^0 state, whereby one electron remains in the conduction band and the other is trapped with the donor ion. The resultant is a neutral centre of higher energy. The two electron condition is known as a D^- or DX^- state, whereby both possible

electrons are trapped with the donor ion. This leaves a single negatively charged entity. More information on these states will be related, but an important point needs emphasising.

At first sight, the two-trapped electrons should have a significant electrostatic repulsion on each other resulting in a positive interaction energy - the so-called Hubbard U. One consequence is that the first electron would be much more tightly bound than the second - a factor of about 20 difference. However if, on the arrival of the second electron, the lattice relaxes, then the binding energy gained from this relaxation, may exceed the repulsive Coulomb energy acting on the first electron. The relaxed crystal is conferring on the electron a local effective negative Hubbard U.

Further experiment has suggested a 'pathway' between the possible states for the ultimate entrapment of two electrons in the DX^- state. Clearly, more than just the possible energy changes of the electrons must be considered. As there is an accompanying change in the elastic potential energy around the local chemical bonding, in other words relaxation, then this transition must also be included. This double occurrence is best visualized using a configuration coordinate diagram, where the total energy of the electron and the lattice is plotted against a configuration coordinate (a measure of lattice distortion). A typical example is shown in fig. 1.4.3.[38,40]

The various confining potentials of the donor states can be represented parabolically on this diagram. Therefore, if a transition, including lattice relaxation, is made from one parabolic state to another, then both the required electronic and configuration energy changes can be seen immediately. There are actually four states shown in this figure. The purple-coloured state is thought to be involved in photo-ionisation

processes, but it is the other three, which are described above, that are important in this work.

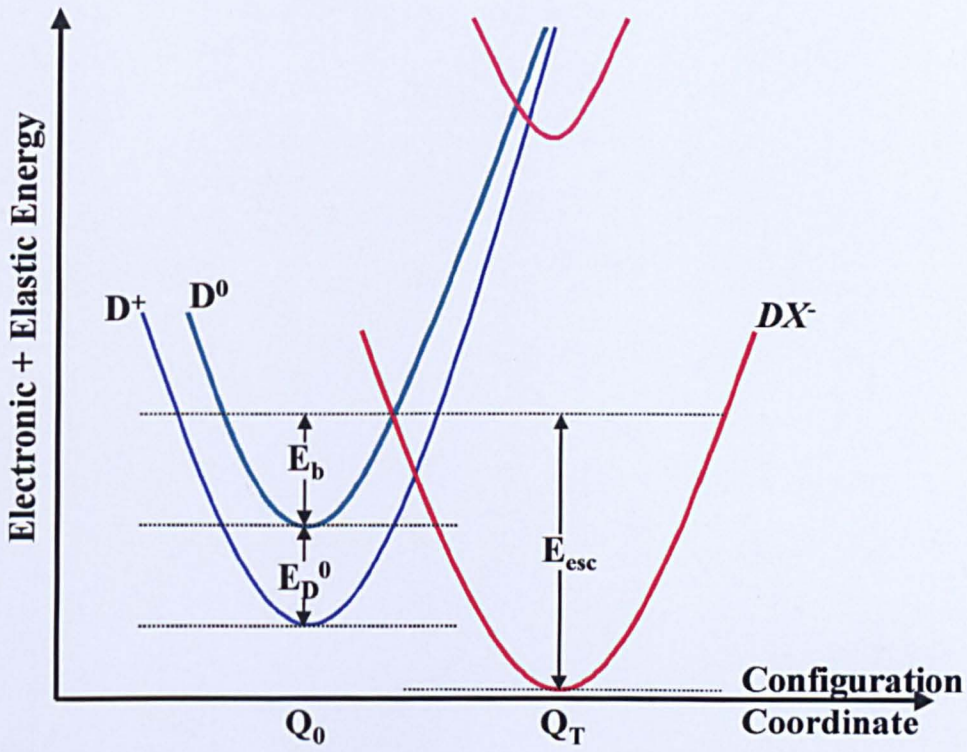


Figure 1.4.3 A typical configuration-coordinate diagram of the 4 known electronic states of a given donor centre. The blue parabola represents the D^+ state; green- the D^0 state; red- the DX^- state; and purple represents a state used in photo-ionization. See the text for an explanation.

The two parabolic states with configuration coordinate Q_0 are those associated with the substitutional configurations. The lower energy state at Q_T represents the distorted *DX* configuration (as indeed does the upper one), which can trap two electrons (DX^-). The energy at this parabola minimum is twice that of the average binding energy of each trapped electron. A reference energy, such as the conduction band-edge, could be shown, but that can vary with *x*.

Thermally induced transitions to the DX^- state are assumed, from experimental evidence, to involve all three of these lower energy states.

A transition can commence with the D^+ state (blue), where both electrons are in the conduction band, hence only the positive donor ion remains in the configuration. If one electron is trapped, then the excited D^0 state (green) is attained, with a gain of potential energy of E_D^0. Note that no relaxation, or a change of configuration coordinate, of the lattice is yet necessary. To trap the other electron, relaxation is required, with a further input of energy. This additional energy is implied (but also experimentally confirmed) from the overlap of the two parabolae representing the two states. (This observation has an analogy, commonly found in many exothermic chemical reactions. Here, an input of energy (the energy of activation) is initially required to start the reaction, resulting in a rearrangement of the potential energy of the chemical bonds, which in turn releases more energy.)

Therefore, there is a maximum thermal potential energy barrier, $E_D^0+E_b$, for the electrons to overcome in order to enter a DX^- state, via a D^0. For an electron to escape, it would require a different energy input, E_{esc}. As stated above and seen in fig. 1.4.1, if the mole fraction (x) of Al is varied, then so do the energies of these various states relative to the conduction band-edge. At large x, a stronger binding energy exists for electrons in the DX^- state hence, a large thermal energy input (E_{esc}) would be required to empty these states - perhaps even beyond that which can be readily supplied at room temperature. This would have consequences on device conductivity at room temperature. Therefore, increasing x, just to enlarge the spacer-channel band-edge offset to result in more channel-confined electrons, may not be an option. (There are, of course, other reasons for not effecting this increase, as related in section 1.3, concerning elastic strain collapse at this interface.)

However, if x is about 0.3, as in this work, one can observe experimentally the effect of the DX^- centres on device parameters, as the thermal energy is varied. At a given temperature, there is enough thermal energy per electron ($E_D^0+E_b$) to enable the DX^- state to trap a significant electron density. At some higher temperature, the electrons can escape again to the conduction band by overcoming the E_{esc} barrier.

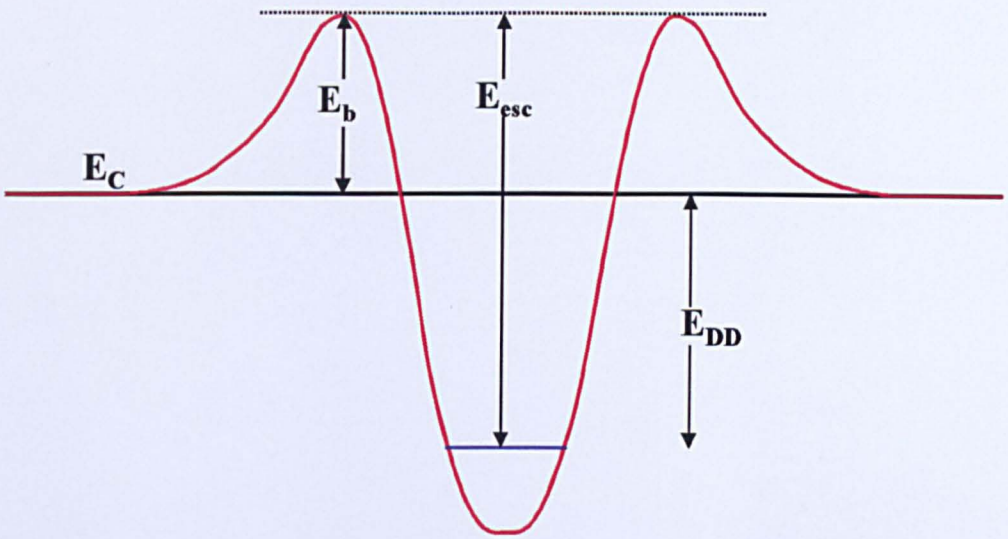


Figure 1.4.4 A schematic representation of the distortion of the conduction band-edge in the vicinity of a *DX* centre.[16] E_{DD} represents the average energy of the two trapped electrons relative to the conduction band-edge (E_C) and, if enough occupied *DX* centres of this type exist, the Fermi level can pin on this blue line. If the Al mole fraction (x) is 0.3, then the effective E_{DD} is about 70meV.

If one starts from a non-equilibrium state with the *DX* centres unoccupied, then as the temperature is increased from a low value, the changing electronic characteristics of a device can be observed,[14 - 16] as the DX^- centres first trap then release their electrons. Fig. 1.4.4 shows a schematic representation of the distortion of the conduction band-edge in the vicinity of a *DX* centre.[16] This is derived from the information presented both by experiment and in fig. 1.4.3. In this figure, it is assumed that the electrons are initially in the D^0 state, that is, around the

conduction band-edge, hence only an energy E_b is required to enter the DX^- state. E_{DD} represents the average energy of the two trapped electrons relative to the conduction band-edge. If enough occupied *DX* centres of this type exist, the Fermi level can pin on this blue line. If the Al mole fraction (x) is 0.3, then a characteristic value for E_{DD} is about 70meV. More about the nature of *DX* centres will be discussed in section 2.3, where the process of 'bias cooling', which can alter their occupancy, is introduced. For detailed summaries on this subject, the reader is directed to the denoted literature.[38-40]

1.5 References

1. Ando T., Fowler A.B., Stern F., Review of Modern Physics, **54**, 437, (1982).
2. Weisbuch C., Vinter B, Quantum Semiconductor Devices, Academic Press, pg.101, (1991).
3. Anderson T.G., Chen Z.G., Kulakovskii V.D., Uddin A., Vallin J.T., Applied Physics Letters, **51**, 752, (1987).
4. Marsh J.H., Applied Physics Letters, **41**, 732, (1982).
5. Cullis A.G., Pidduck A.J., Emeny M.T., Journal of Crystal Growth, **158**, 15, (1996).
6. Littlejohn M.A., Hauser J.R., Glisson T.H., Ferry D.K., Harrison J.W., Solid State Electronics, **21**, 107, (1978).
7. Luo J.K., Ohno H., Matsuzaki K., Hasegawa H., Japanese Journal of Applied Physics, **27**, 1831, (1988).
8. Basu P.K., Nag B.R., Applied Physics Letters, **43**, 689, (1983).
9. Bhattacharyya K., Orwa J.O., Goodnick S.M., Journal of Applied Physics, **73**, 4396, (1993).
10. Skromme B.J., Tamargo M.C., Turco F.S., Shibli S.M., Bonner W.A., Nahory R.E., Institute of Physics Conference Series, No. **96**, 205, (1989).
11. Toyoshima H., Onda K., Mizuki E., Samoto N., Kuzuhara M., Itoh T., Okamoto A., Anan T., Ichikashi T., Journal of Applied Physics, **69**, 3941, (1991).
12. Bhattacharya P., Semiconductor Optoelectronic Devices, Prentice-Hall, (1997).
13. Neamen D.A., Semiconductor Physics and Devices- Basic Principles, Richard D Irwin, (1992).

14. Long A.R., Davies J.H., Kinsler M., Vallis S., Holland M.C., Semiconductor Science and Technology, **8**, 1581, (1993).

15. Skuras E., Holland M.C., Barton C.J., Davies J.H., Long A.R., Semiconductor Science and Technology, **10**, 922, (1995).

16. Davies J.H., The Physics of Low-Dimensional Semiconductors, Cambridge University Press, (1998).

17. See for example van der Burgt M., Karavolas V.C., Peeters F.M., Singleton J., Nicholas R.J., Herlach F., Harris J.J., Van Hove M., Borghs G., Physical Review B, **52**, 12218, (1995).

18. Sze S.M., Physics of Semiconductor Devices, 2nd Edition, Wiley, (1982).

19. Shur M., Physics of Semiconductor Devices, Prentice-Hall, (1990).

20. Singh J., Physics of Semiconductors and Their Heterostructures, McGraw-Hill, (1993).

21. Matthews J.W., Blakeslee A.E., Journal of Crystal Growth, **27**, 118, (1974).

22. Jaffe M., Singh J, Journal of Applied Physics, **65**, 329, (1989).

23. Kim K.W., Littlejohn M.A., I.E.E.E. Transactions on Electron Devices, **38**, 1737, (1991).

24. See for example Li G., Babinski A., Jagadish C., Applied Physics Letters, **70**, 3582, (1997).

25. Lang D.V., Logan R.A., Physical Review Letters, **39**, 635, (1977).

26. Lang D.V., Logan R.A., Jaros M., Physical Review B, **19**, 1015, (1979).

27. Mizuta M., Tachikawa H., Kukimoto H., Minomura S., Japanese Journal of Applied Physics, **24**, L143, (1985).

28. Theis T.N., Mooney P.M., Wright S.L., Physical Review Letters, **60**, 361, (1988).

29. Theis T.N., Mooney P.M., Parker B.D., Journal of Electronic Materials, **20**, 35, (1991).

30. Mooney P.M., Theis T.N., Wright S.L., Applied Physics Letters, **53**, 2546, (1988).

31. Baba T., Mizuta M., Fujisawa T., Yoshino J., Kukimoto H., Japanese Journal of Applied Physics, **28**, L891, (1989).

32. Brunthaler G., Köhler K., Applied Physics Letters, **57**, 2225, (1990).

33. Calleja E., Mooney P.M., Theis T.N., Wright S.L., Applied Physics Letters, **56**, 2102, (1990).

34. Calleja E., Garcia F., Gomez A., Muñoz E., Mooney P.M., Morgan T.N., Wright S.L., Applied Physics Letters, **56**, 934, (1990).

35. Morgan T.N., Journal of Electronic Materials, **20**, 63, (1991).

36. Chadi D.J., Chang K.J., Physical Review B, **39**, 10063, (1989).

37. Chadi D.J., Chang K.J., Physical Review Letters, **61**, 873, (1988).

38. Mooney P.M., Semiconductor Science and Technology, **6**, B1, (1991).

39. Lang D.V., Deep Centers in Semiconductors- A State of the Art Approach, 2nd Edition, Chapter 7, pg 591; Pandelides S.T. (ed.), Gordon and Breach, (1992).

40. Mooney P.M., Deep Centers in Semiconductors- A State of the Art Approach, 2nd Edition, Chapter 7, pg 643; Pandelides S.T. (ed.), Gordon and Breach, (1992).

CHAPTER 2

ELECTROSTATIC AND TRANSPORT MODELS

2.1 An Electrostatic Model

The ability to make predictions with regard to the various parameters, which characterise a semiconductor layer system, is very beneficial for device modelling. This applies whether the device will be discrete or part of an integrated circuit. Examples of these parameters include the channel's charge carrier density, the minimum doping density, which will achieve that density and the threshold or cut-off gate voltage.

A relatively simple model based on electrostatic principles was developed by Long et al (1993).[1] Analytic expressions were formulated, using only the Laws of Gauss and Poisson, for all of the above-mentioned parameters and some others. Although some necessary approximations and assumptions were included, the model did,

however, render unnecessary more complicated self-consistent calculations to evaluate these parameters. The model was constructed particularly with GaAs/AlGaAs systems in mind, but could be applied to any similar system in principle. The original paper[1] concentrated on slab-doped systems, where the doping is distributed widely over the upper AlGaAs layer. A second paper by Skuras et al (1995)[2] extended the principles to a delta-doped GaAs/AlGaAs system.

One of the most important assumptions incorporated in both papers is that the Fermi level is pinned by sufficient donor electrons occupying the *DX* centres sited in the doping layer, see section 1.4. As explained there, electrons trapped in such centres will occupy deep donor states about 70meV ($Al_{0.3}Ga_{0.7}As$) below the local conduction band. The pinning is developed at room temperature, where electron exchange with *DX* centres, although slow, can occur and the system can equilibrate. Davies (1998)[3] develops these electrostatic principles, without such pinning. In addition to the development of the model, considerable experimental work was carried out to compare the effect of full and empty *DX* centres on the threshold gate voltage. A method of cooling under bias was developed to change the occupation, but more detail on this will be given in section 2.4.

The pseudomorphic systems vary from the lattice-matched (GaAs/AlGaAs) systems, in having both strain induced within the channel and also an extra band-edge energy offset between the channel and the substrate. However, at the commencement of the work, it was thought that the model might be extended to include these systems.

The derivation of the relevant equations requires an analysis of the distribution of charge and its associated electric fields. The energy changes of the conduction band-edge over this charge distribution are

also included, with, of course, the deep-donor energy providing a useful energy reference. In what follows, a derivation of only one of the equations will be outlined, culminating in the expression for the channel carrier density. The remaining relations will merely be stated, as their derivations evolve in much the same manner. Fig. 2.1.1 should offer the reader a guide during the derivation.

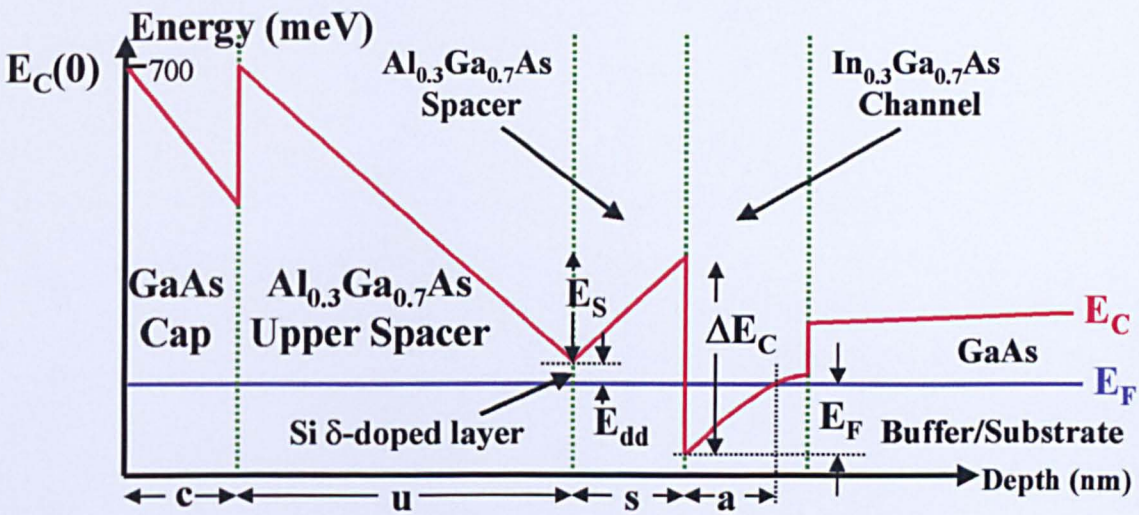


Figure 2.1.1 Conduction band-edge energy profile, See the text for an explanation of the symbols.

The zero of energy is placed on the Fermi level and the lowest value of energy (E_C) is defined as $-E_F$. The main assumptions are:

1. As stated earlier, the Fermi level is pinned on the deep-donor energy states (E_{dd}), and the electrons cannot be removed at low temperature by the application of a gate bias.
2. The energy barrier of the spacer is insufficiently high to prevent tunnelling of electrons from the channel to the doping layer, even at low temperatures. In this way, the Fermi level can equate across the whole layer system.

3. To this end, it is further assumed that the Fermi level will also pin on the large electron occupation trapped in the mid-band-gap surface states of the GaAs cap. A Schottky barrier ($E_C(0)$) of about 700meV is produced on deposition of a gate metal.
4. In response to a repulsive gate field, the energy barrier between the cap and the spacer is sufficiently high to prevent the penetration of electrons from the surface, and hence, the formation of holes there.
5. Any electron charge in the channel is degenerate, that is, there is only single subband occupancy.
6. The expression for the Fermi level (E_F), derived from the Variational Theory[4] (K=1.6), or the Triangular Well Approximation (K=1.85) is (in a 'clean' system free from defects and impurities):

$$E_F = K\left[\left(\frac{e^2\hbar}{\varepsilon_c\varepsilon_0}\right)\frac{1}{n_{2D}}\right]^{\frac{1}{3}} n_{2D} + \frac{\pi\hbar^2}{m^*} n_{2D} \qquad (2.1.1)$$

where e is the electronic charge; $\hbar$ is Planck's constant/2π; ε_c is the channel's relative permittivity; ε_0 is the free-space permittivity; m^* is the effective mass of an electron and n_{2D} is the 2DEG density. The first term quantifies the energy difference between the conduction band-edge and the base of the first subband and the second term the energy difference between the latter and the Fermi level. This can be approximated by the expression (merely stated in ref. 1):

$$E_F = \frac{n_{2D}e^2a}{\varepsilon_c\varepsilon_0} \qquad (2.1.2)$$

where a is the effective channel width. This approximation can be re-arranged to:

$$n_{2D} = \frac{E_F\varepsilon_c\varepsilon_0}{e^2a} \qquad (2.1.3)$$

Applying Gauss' Law across the spacer-channel boundary gives:

$$\int_S \mathbf{F}_c.d\mathbf{A} = \frac{Q}{\varepsilon_c \varepsilon_0} \Rightarrow \varepsilon_c |\mathbf{F}_c| = \frac{Q}{\varepsilon_0 A} = \frac{n_{2D} e}{\varepsilon_0} \tag{2.1.4}$$

where F_c is the electric field on the side of the channel; dA is an area element and Q is the total charge in the channel. Equation 2.1.4 can also be equated with the field ($\mathbf{F}_s$) at the spacer side of the junction in the following way:

$$\varepsilon_c |\mathbf{F}_c| = \varepsilon_s |\mathbf{F}_s| = \frac{n_{2D} e}{\varepsilon_0} \tag{2.1.5}$$

where ε_s is the spacer's relative permittivity. With reference to fig. 2.1.1, it can be seen that:

$$E(c+u) = E_{dd} = -E_F + \Delta E_c - E_s = -E_F + \Delta E_c - e|\mathbf{F}_s|s \tag{2.1.6}$$

where ΔE_c is the conduction band-edge offset between the channel and the spacer; E_s is the change in the band-edge energy across the spacer and s is the spacer width. Substituting equation 2.1.2 for E_F and equation 2.1.5 for $|F_s|$ into equation 2.1.6 gives:

$$E_{dd} = -\left(\frac{n_{2D} e^2 a}{\varepsilon_0 \varepsilon_c}\right) + \Delta E_c - \left(\frac{n_{2D} e^2 s}{\varepsilon_0 \varepsilon_s}\right) \tag{2.1.7}$$

$$\Rightarrow E_{dd} - \Delta E_c = -\frac{n_{2D} e^2}{\varepsilon_0}\left[\frac{a}{\varepsilon_c} + \frac{s}{\varepsilon_s}\right] \tag{2.1.8}$$

Re-arranging equation 2.1.8 in terms of the density of the 2DEG gives:

$$n_{2D} = \frac{(\Delta E_c - E_{dd})\varepsilon_0}{\left(\frac{a}{\varepsilon_{channel}} + \frac{s}{\varepsilon_{spacer}}\right) e^2} \tag{2.1.9}$$

This equation shows that $n_{2D} \sim 1/s$ for large s, but for $s \leq a$, n_{2D} cannot be increased indefinitely. Also $n_{2D} \sim \Delta E_C - E_{dd}$, that is, it rises with the band offset. The other equations can be derived using similar techniques and are:

$$N_\delta = \frac{n_{2D}\left(\frac{c}{\varepsilon_{cap}} + \frac{u}{\varepsilon_{upperspacer}} + \frac{m}{\varepsilon_{monolayers}} + \frac{s}{\varepsilon_{spacer}} + \frac{a}{\varepsilon_{channel}}\right) + \frac{\varepsilon_0}{e^2}E_c(0)}{\left(\frac{c}{\varepsilon_{cap}} + \frac{u}{\varepsilon_{upperspacer}}\right)} \qquad (2.1.10)$$

where N_δ is the critical doping density – the doping density which is sufficient to charge both the surface states and the channel for this model to be valid; c, u, and m are the widths of the cap, upper spacer and monolayers respectively and ε_j is the relative permittivity of the undernoted (j) layer.

$$V_T^f = -\frac{e}{\varepsilon_0}n_{2D}\left(\frac{c}{\varepsilon_{cap}} + \frac{u}{\varepsilon_{upper\ spacer}} + \frac{m}{\varepsilon_{monolayers}} + \frac{s}{\varepsilon_{spacer}} + \frac{a}{\varepsilon_{channel}}\right) \qquad (2.1.11)$$

where V_T^f is the threshold gate voltage when all the electrons in the donor layer are frozen (f) into the deep-donor centres at low temperature.

$$\frac{C^f}{A} = \varepsilon_0\left(\frac{c}{\varepsilon_{cap}} + \frac{u}{\varepsilon_{upper\ spacer}} + \frac{m}{\varepsilon_{monolayers}} + \frac{s}{\varepsilon_{spacer}} + \frac{a}{\varepsilon_{channel}}\right)^{-1} = e\frac{dn}{dV_G} \qquad (2.1.12)$$

where C^f/A is the 'frozen' capacitance per unit area and dn/dV_G is the channel depletion rate with gate voltage.

2.2 Shubnikov-de Haas and Hall Measurements

2.2.1 Introduction

The measurement systems implemented in this work mainly output the resistance or conductance accessed from the appropriate section of a Hallbar, see chapter 4. These parameters vary with gate voltage and magnetic field. Whilst some of these measurements are useful in their own right, such as the variations in magneto-resistance found in Shubnikov-de Haas (SdH) and Quantum Hall Effect plots, they generally require further manipulation for meaningful analysis. Using the simple principles behind the classical Hall effect, parameters such as the transport Hall mobility and the Hall carrier density can be evaluated.

In what follows, the principles behind the above parameters will be outlined. The use of the quantum Hall Effect and, in particular, the SdH plots will also be summarised. In particular, their limitations will be emphasised.

2.2.2 Shubnikov-de Haas Oscillations and the Quantum Hall Effect

As stated in the introduction, these methods are useful in their own right but can also complement the classical Hall measurements. Both the evolution of the SdH oscillations, see, for example, fig.5.1.1, and the non-linear plateaux of the quantum Hall Effect require a quantum mechanical explanation. The presence of a magnetic field

directed perpendicularly to the 2DEG of a heterostructure alters the normal constant 2-D distribution of the density of states over energy. There is complete quantization of the cyclotron energy of the electrons, and hence, ideally, their density of states. Also required are low temperatures, such that $\omega_c\tau>1$ and $\hbar\omega_c > k_BT$, where ω_c is the cyclotron angular frequency, τ is the mean scattering time, k_B is Boltzmann's constant and T is the temperature.

These cyclotron energy levels are called Landau Levels, and their separation is proportional to the magnitude of the magnetic field. In real systems, however, some broadening of this distribution of the density of states over energy will arise owing to impurities, defects, alloy fluctuations, interface roughness etc. The energy (E_n) of the centre of the n^{th} Landau Level is given by:

$$E_n = (n+1/2)\hbar\omega_c \qquad (2.2.1)$$

where n=0,1,2,3...; and ω_c is equal to eB/m^*, where $\boldsymbol{B}$ is the magnetic field. Clearly, as the magnetic field increases, the energy separations of the levels diverge. These levels and their associated distribution of states will, therefore, pass through the Fermi level in turn until the magnetic field is sufficiently high that the lowest energy level ($\hbar$eB/2m*) can pass through. As the states of occupation pass through the Fermi level, the longitudinal magneto-resistance would be expected to change in an oscillatory manner. This is experimentally observed in the SdH oscillations.

Many attempts have been made to account analytically for these oscillations with Ando et al[5] (1982) and Coleridge et al[6] (1989) making successful contributions. Their expressions apply only at low magnetic field such that $\omega_c\tau$, where τ is the elastic scattering time, is less than or about equal to 1. These are fairly complicated expressions

but they contain information on the effective mass of the electrons and their quantum mobility. The latter can be extracted from the rate of change of the oscillatory amplitude. The equation for the resistivity given by Coleridge et al (1989) is:

$$\rho_{XX} = \rho_0\left[1 - 4\exp\left(\frac{-\pi}{\omega_c\tau_q}\right)\frac{X}{\sinh X}\cos\left(\frac{2\pi E_F}{\hbar\omega_c}\right)\right] \qquad (2.2.2)$$

where ρ_{xx} is the longitudinal magneto-resistance; ρ_0 is the zero field magneto-resistance, which can also be expressed in terms of the transport relaxation time term (τ_{Tr}) using $\rho_0 = m^*/ne^2\tau_{Tr}$; τ_q is the quantum relaxation time; and $X = 2\pi^2 k_B T/\hbar\omega_c$. The information relevant to calculate quantum mobility (equation 2.2.9) can be extracted from equation 2.2.2 with an additional term R_S (equation 2.2.4) included to account for any spin-splitting. The adapted equation is:[7]

$$\frac{\rho_{xx}^{peak}}{\rho_0} = 1 + 4R_S R_T R_\Gamma \qquad (2.2.3)$$

where

$$R_S = \cos\left(\frac{\pi g m^*}{2m_0}\right) \qquad (2.2.4)$$

$$R_T = \frac{X}{\sinh(X)} \qquad (2.2.5)$$

$$\text{and } R_\Gamma = \exp\left(\frac{-\pi}{\omega_c\tau_q}\right) \qquad (2.2.6)$$

with g the gyromagnetic ratio; m_0 the free electron mass and ρ_{xx}^{peak} the amplitude of the low-field oscillations.

This implies that a "Dingle" plot of:

$$\mathrm{Ln}\left[\left(\left(\frac{\rho_{xx}^{peak}}{\rho_0}\right)-1\right)/4R_S R_T\right] \tag{2.2.7}$$

versus *1/B* would yield a gradient (s) of:

$$s=-\frac{\pi m^*}{e\tau_q} \tag{2.2.8}$$

from which τ_q can be extracted. Therefore, the quantum or 'single particle' relaxation time (τ_q) can be inferred from the exponential decay of a SdH plot with decreasing magnetic field based on the above expressions, provided that the field is not too high ($\omega_c\tau_q<1$). Every scattering event is important in determining τ_q, whereas the transport scattering time, and hence the measured transport mobility, lends more weight to large-angle scattering. The corresponding quantum mobility (μ_q) is simply given by:

$$\mu_q=\frac{e\tau_q}{m^*} \tag{2.2.9}$$

A modified Dingle plot[7] is applicable when the quadratic relation between the magneto-resistance and the density of states at the Fermi level is taken into account (Only given for information as it is not used here. See also page 56 for further descriptions on magneto-oscillations):

$$\sqrt{\frac{\rho_{xx}^{peak}}{\rho_0}}=1+2R_S R_T R_\Gamma \tag{2.2.10}$$

The relevant plot, in this case, would be:

$$\mathrm{Ln}\left[\left(\sqrt{\left(\frac{\rho_{xx}^{peak}}{\rho_0}\right)}-1\right)/2R_S R_T\right] \tag{2.2.11}$$

versus 1/B. This relation is valid over a greater range of magnetic field than the normal Dingle plot. Other expressions are available for very high fields.

From the reciprocal 'period' ($\Delta(1/B)$) of the SdH oscillations, the carrier density (n) can be inferred from the following expression:

$$n = \frac{2e/h}{\Delta(1/B)} \tag{2.2.12}$$

where *h* is Planck's constant. Alternatively, the Fourier transform of the SdH plot in *1/B* will provide the peak, or most common, frequency (ν_{pk}). The presence of occupied subbands at higher energy is indicated by the presence of additional peaks on such plots. The quantum mobility can also be inferred from the width of these peaks. The raw SdH data itself can indicate multiple subband occupancy through the appearance of interference 'beats'. A rising (B^2) background to the envelope suggests the presence of a parallel conductive path.[8] Equation 2.2.12 can be modified to:

$$n = \frac{2e\nu_{pk}}{h} \tag{2.2.13}$$

This was the expression used to determine carrier density. An expression for mobility can be deduced from the longitudinal resistance at zero magnetic field (R_{xx} at B=0T):

$$R_{XX} = \rho_{XX}\frac{L_X}{L_Y} = \frac{L_X}{\sigma_{XX}\,L_Y} = \frac{L_X}{L_Y}\frac{1}{ne\mu} \tag{2.2.14}$$

where σ_{xx} is the longitudinal conductivity; L_j is the length of the measured pathway in the *j* direction, μ is the mobility and *n* is evaluated using equation 2.2.13. Therefore, the mobility can be expressed by:

$$\mu = \frac{L_X}{L_Y} \frac{1}{R_{XX}(B = 0T)ne} \tag{2.2.15}$$

and this was the equation used to calculate mobility from the SdH plots.

2.2.3 Classical Hall Measurement

Three resistance measurements, derived from a standard 8-terminal Hallbar, see fig. 3.2.4c, are required for the evaluation of a Hall carrier density and its associated mobility. Two of these are required for theoretical reasons and the other is necessary to counteract a practical problem. The two measurements satisfying the theory are the longitudinal resistance (R_{xx}) measured at zero magnetic field along a fixed length (L_X) (300μm) of the Hallbar, and the Hall resistance, which is measured across a fixed width (L_Y) (100μm) of the Hallbar. The latter is acquired in the presence of a constant, low and perpendicular magnetic field. The field must be low enough not to produce a detectable quantum Hall Effect, that is it must be classical. The SdH plot is one method of ensuring that the quantum limit (the oscillations) has not been attained. The third, and final, measurement is another Hall resistance acquired at a zero magnetic field. This is effected to counteract factors existing in real samples, which cause electrons to preferentially favour one side of the Hallbar to the other when directed along it. This bias will cause an offset potential difference to appear across the Hallbar, which in the absence of a magnetic field, should theoretically not exist. It, therefore, must be subtracted from the other Hall resistance.

Each resistance measurement is acquired whilst a gate voltage is swept, depleting the Hallbar of most of its electrons. In this way, the

effect of carrier density on mobility can be determined. The three gate-voltage-dependent data sets acquired are **a)** R_{xx} (B = 0T); **b)** R_{xy} (B = 0T); and **c)** R_{xy} (B ~ 1T).

The formulae used to calculate the Hall mobility (μ_H) and the carrier density (n_H) can be derived from the expression for current density (j) as follows:

$$j = n_H e|v_B| = \frac{n_H e|F|}{|B|} = \sigma|F| \tag{2.2.16}$$

where $|v_B|$ is the mean drift speed of electrons in a magnetic field; $|F|$ is the magnitude of the electric field and σ is the conductivity.

$$\Rightarrow \sigma = \frac{n_H e}{|B|} \Rightarrow \rho = \frac{|B|}{n_H e} = R_{xy} \tag{2.2.17}$$

where ρ is the resistivity. Re-arranging equation 2.2.17 in terms of the carrier density gives:

$$n_H = \frac{|B|}{eR_{xy}} \equiv \frac{|B|}{e\Delta R_{xy}} \tag{2.2.18}$$

where the Δ sign, and implied subtraction, accounts for the experimental offset discussed earlier. The Hall mobility (μ_H) can be derived from the corresponding 'SdH' equation 2.2.15:

$$\mu = \frac{L_X}{L_Y} \frac{1}{R_{XX}(B = 0T) n e} \quad \text{but with} \quad n = \frac{|B|}{e\Delta R_{XY}} \tag{2.2.19}$$

$$\Rightarrow \mu = \frac{L_X}{L_Y} \frac{\Delta R_{xy}}{|B| R_{XX}(B = 0T)} \tag{2.2.20}$$

The Limitations of the Hall Measurements

Equations 2.2.18 and 2.2.20 can only be applied if one subband is occupied in the channel. In addition, they are not applicable when two or

more conductive paths in the layer system are detected. The former is signified by beats in the SdH oscillations, while the latter generally produces a rising background to the oscillatory envelope. Algebraically, it can be shown that if two or more conductive paths exist then the average magneto-resistance develops a quadratic (B^2) term with magnetic field.

van der Burgt et al (1995)[8] illustrate this effect when they create additional conductive pathways by emptying the *DX* centres in their pHEMT devices using illumination. The excess electrons occupy the shallow states in the donor layer, where electrical excitation, supplied between the source and drain, produces a low mobility current. This paper also shows the effect of multiple subband occupation in the channel on the Hall mobility. The conclusion emerging is one of a carrier density and mobility which is weighted according to the carrier density and mobility of each subband (i) as follows:

$$n_H \mu_H = \sum_i n_i \mu_i \qquad (2.2.21)$$

with

$$n_H = \frac{\left(\sum_i n_i \mu_i\right)^2}{\sum_i n_i \mu_i^2} \qquad (2.2.22)$$

where n_H and μ_H are the weighted (Hall) carrier density and mobility, respectively and n_i and μ_i are the carrier density and the mobility, respectively, of the i^{th} subband, measured from SdH data. Clearly, if only one subband is occupied, then $n_H = n_1$ and $\mu_H = \mu_1$. The above equations need not be restricted to different subbands but can also be used on parallel conduction pathways.

2.3 The Effect of Correlation in the Doping Layer

2.3.1 Introduction

In this section, the influence of the electrostatic correlations between electrons in the donor layer will be discussed. Their chief effect is in the reduction of scattering of the channel electrons by the ionized donors in the donor or doping layer. The *DX* and associated centres play a crucial part[9,10] and their characteristics have already been discussed in section 1.4.

2.3.2 The Establishment of Donor Layer Correlation

By the implementation of certain procedures, which will be outlined later, the net charge density within the doping layer can be greatly adjusted. Studies of this kind[11,12,14] have helped to re-affirm the applicability of the negative-U, DX^- model of a donor.[9,10] If the *DX0* model is implemented, large variations in the net charge density (ρ_0) of the doping layer would be expected to produce commensurate variations in the 2DEG mobility. In this situation of d^+, d^0 and *DX0* donor types, only the substitutional, hydrogenic ionised donors would be expected to produce a scattering field. Therefore, if ρ_0 is adjusted, then the scattering field of the d^+ donors would vary similarly. Ignoring the possible presence of correlation (in this case a smoothing of the donor layer potential) by occupation, this was not observed.[11]

The implementation of a *DX* model for some of the donors in place of the *DX0* model would lend credence to the higher mobility observed. If two electrons could be trapped in these meta-stable sites then, at low temperatures, their net charge (-1) is effectively fixed and would be able to counteract somewhat, the field due to the d^+ donors. The higher 2DEG mobility observed is more in accord with this reduced scattering field emanating from the doping layer. An electrostatic model,[1] formulated to calculate various FET parameters, is more accurate when the *DX* construction is incorporated and the Fermi energy is pinned to its characteristic energy.

With the *DX* model now included, the subtler but still significant changes in mobility arising from the donor layer correlation becomes more explicable, although correlation can also occur in a single electron occupation mode. It would be expected that correlation would be at a maximum when the number of occupied deep *DX* centres and empty shallow d^+ donors are equal. Alternatively, no correlation would be expected when neither the *DX* nor the d^+ centres exist. A useful ratio (η) for expressing this variable correlation is mentioned by Buks et al.[11] It relates the difference between the charge density of the above two situations (N_{DX^-} and N_{d^+} respectively) to their combined total density (N):

$$\eta = \frac{N_{d^+} - N_{DX^-}}{N} \qquad (2.3.1)$$

This ratio can range theoretically between the two extremes of -1 and $+1$, when all the donors are *DX* like and d^+ like respectively. The situation of maximum correlation would imply that N_{d^+} and N_{DX^-} be approximately equal and hence the ratio η is zero.

In addition to this Coulombic correlation interaction, there would also exist a much smaller correlation due to any non-random distribution of the Si donors. This addition would have a relatively negligible effect on mobility. Subsequent to a strong negative bias cool – a method described later in section 2.3.3 - the occupation of the *DX* donors would be expected to be slight and hence η would tend to a value of +1. Application of a less negative bias cool would allow η to tend towards a value of zero. The onset of increased gate leakage, in response to a high positive gate voltage, will, in practice, prevent the regime of $-1 < \eta < 0$ being attained. This regime is also theoretically unattainable through unphysical solutions of Poisson's equation, and the electrons occupying channel and surface states need to come from somewhere.

Experiments[11] on GaAs/AlGaAs systems have shown that η can get very close to zero (0.08) or one (0.94) under the appropriate conditions of high positive and high negative bias cooling respectively. The mobility of the channel electrons also varies generally as expected. Experimentally, η can be determined from the net 2D donor charge density (ρ_0), using the capacitance per unit area between the gate and the δ-doped layer (C_δ), the depletion gate voltage (V_D) after a normal cool and the Schottky barrier potential (V_0) using the following relation:

$$\rho_0 = C_\delta (V_0 - V_D) \tag{2.3.2}$$

where C_δ is determined from:

$$C_\delta^{-1} = \left(\frac{e\varepsilon_0 \varepsilon_{channel}}{z_2 - z_1} \right)^{-1} + \left(\frac{e\varepsilon_0 \varepsilon_{spacer}}{z_1 - d} \right)^{-1} \tag{2.3.3}$$

where ε_j is the relative permittivity of the undernoted layer j; z_2 is the gate-channel distance; z_1 is the cap-channel distance and d is the spacer thickness.

Regardless of the magnitude of the bias cooling in the *DX* model, the total density of the charged ions (N) is constant, therefore, any mobility changes can only be attributed to changes in the doping layer correlation. With increasing correlation, the change in mobility (μ) with carrier density (n) departs from the typical $\mu \propto n^{3/2}$ trend, derived theoretically for remote ionised impurity (RII) scattering (see section 2.4), to a more linear relationship. Remote ionised impurity scattering arises from the scalar potential field generated by the charged donors. Generally, theoretical derivations for RII scattering do not include correlation.

However, Buks et al[11] derive a 'correlator factor' ($\chi(\xi,\nu_q)$) which can be incorporated into a general expression for uncorrelated RII scattering. For this factor, the smaller fluctuations (about inter-atomic lengths) were averaged out and only those with a length scale greater than or equal to the spacer width were considered. It is given by:

$$\chi(\xi,\nu_q)=\frac{1+\nu_q\xi^2}{(1+\nu_q\xi)^2} \tag{2.3.4}$$

where

$$\xi=1-\eta^2 \tag{2.3.5}$$

and

$$\nu_q=\frac{n_0 e^2}{k_B T_f}Q_q \tag{2.3.6}$$

where n_0 is the average impurity density and T_f is the 'freeze-out' temperature and:

$$Q_q=\frac{1}{2\varepsilon_{spacer}\varepsilon_{channel}\,q}\left[1-\exp(-2d)\right] \tag{2.3.7}$$

where q is the scattered vector.

Note from equations 2.3.4 and 2.3.5 that if $\eta = \pm 1$, then $\xi = 0$ and hence $\chi = 1$ and no correlation is predicted. (If $\eta = 1$, the density of occupied *DX* centres is zero.)

2.3.3 Experimental Methods for Varying Correlation

2.3.3.1 Bias Cooling

This was the method of choice here as in other work[1,2,11,12] because it is convenient to implement on gated layers. Before the sample is cooled a potential is applied to the gate. The resulting field is effective at the level of the channel and doping layer in influencing the average distribution with depth of the electrons. In other words, a positive field will enhance the probability of electrons occupying potential minima closer to the surface, whereas a negative field will discourage this. Therefore, application of the former would be expected to encourage the electron occupation of the *DX* centres in the doping layer, while the latter would lessen this probability. The electrons are held at the desired depth distribution while the *DX* centres are still thermally active. The sample is then cooled slowly to below the temperature, and hence thermal energy, at which the centres become inactive (the freezing temperature T_f). Below this temperature, (*DX* centre type dependent) electrons can neither access nor escape from the metastable *DX* centres when exposed to a subsequent gate field applied during measurement. By this means, occupation of, and the correlations developed by the *DX* centres can be relatively fine-tuned.

The main disadvantage of the bias cooling technique lies in low tolerance of the device layers to a large positive gate bias. Gate leakage will generally develop as the conduction band distorts under its influence, although using deeper channels should limit the onset of this current. This factor clearly limits the degree of high correlation obtainable, but another method described below can counteract this limitation.

2.3.3.2 Other Methods Used for Correlation Adjustment

Assuming the negative-U *DX* model, the two other main procedures for altering the doping layer correlation, apart from thermal excitation, are exposure to illumination and high pressure freeze-out (HPFO). The former is frequently implemented to excite electrons out of the deep donor states, hence reducing the intrinsic correlations of the doping layer. The mobility of the channel will thus fall assuming no electronic screening by electrons occupying the shallow states of the doping layer. The average wavelength (usually provided by an L.E.D.) must be carefully chosen to provide sufficient energy to remove the electrons from their deep-level state to the conduction band. This energy will, of course depend on the *DX* type, see section 1.4. Too much energy could place electrons into other conduction band minima, or even promote electron-hole generation in, for example, the channel layer[13] although they will recombine. In contrast, too little energy ($<E_{dd}$) will, of course, be insufficient to ionise the deep donor centres. Even a suitable energy can depopulate acceptor or impurity states, which may mislead somewhat. Although this method is used to test the effect of

correlations on mobility, it is perhaps more often implemented in varying the relative charge densities of the doping and channel layers.[8]

The latter method, (HPFO) is a far less common procedure for adjusting *DX* centre occupation.[14] As with the gate potential in bias cooling the pressure must be applied to the layers continually during the cooling stage for exactly the same reasons. This is achieved by mounting the sample in a Cu-Be high-pressure chamber with a helium gas compressor. Typical pressures can range to about 10kbar, thus allowing some tuning of the *DX* centre occupation. The pressure has the effect of distorting the conduction band, which mainly affects the channel layer here by forcing electrons out of this region. These electrons will then seek other minima of the band-edge such as the doping layer where the *DX* centres are sited. This pressure enhances the *DX* occupation, whilst depleting the channel states. Therefore, it has the same effect as a positive bias cool, but is much more effective and wide-ranging. Of course, the pressure must be released at low temperatures, otherwise other parameters such as strain, effective mass etc., will be affected.

Other authors[15] have also used pressure, but the work of Dmowski et al is particularly relevant as it deals with pseudomorphic layers. They also apply infra-red radiation to accomplish the complimentary effect of *DX* centre depopulation. They measure the effect of the doping layer correlations by study of the relative variations in quantum and transport mobility as the population of the *DX* centres changes. By combining both the methods of pressure and illumination, they could vary the channel density by about 100%, compared with the few percent possible with bias cooling. They found that the quantum

mobility was relatively much more sensitive to changes in the correlations of the doping layer than the transport mobility.

2.4 Scattering Processes

2.4.1 Introduction

The mobility of electrons in these pHEMT and other layers is fundamentally limited by the frequency of electron collisions or interactions. These collisions can be elastic or inelastic and their frequency can be expressed through the momentum relaxation time (τ), - the average time between interactions or collisions, weighted to large-angle processes. At higher temperatures, the inelastic scattering component dominates the mobility, where the electrons interact with the quantum thermal vibrations (phonons) of the lattice. However, in this section, the emphasis will be placed on the significant elastic scattering mechanisms believed to exist, whose effect is best studied at low temperature.

We will concentrate on the theoretical formulations for the remote ionised impurity (RII) and random alloy scattering factors, as a determination of their characteristics is believed crucial to much of this analysis. Other components, such as interface roughness and cluster scattering, while important, will not be discussed at this stage. The significance of the different mechanisms can best be determined by a comparison of the overall experimental mobility with that derived for each individual component. As stated in section 2.3, a doping layer correlation factor can, in principle, be incorporated into an expression for RII scattering. The approach here, however, is to acquire experimentally, information on the correlation through bias cooling and

subsequently add this to the theoretical derivation where correlations are absent.

The scattering of an electron involves its transition from one state to another, which may also similarly affect the perturbing potential causing the scattering. A time-dependent approach involves study of the system's time evolution as the perturbation is encountered. This approach is perhaps better for inelastic processes, therefore a 'stationary-state' approach will be adopted for this work. It should be stressed that any given derivation should not be considered as "correct", because other approaches can be equally valid. In all cases, assumptions and simplifications must be made to keep the mathematics manageable, and to obtain analytical expressions, which can be solved numerically. The methods and resulting expressions that are most successful and widely used are those which implement these conditions.

There is quite a long list of possible scattering sources in layers of this type and include those arising from:

1. The potential existing in the 2DEG originating from the ionised donors, remotely sited on the other side of the spacer - RII scattering.
2. The potential created from the random distribution of the ternary alloy atoms in the channel layer itself – random alloy scattering.
3. As above, but with a less random distribution – cluster scattering.
4. The intimate interaction of the channel wavefunction with possible atomic disorder at the channel-spacer boundary or other interfaces.
5. Local interactions with background impurities – neutral or ionised.
6. Local interaction with the lattice vibrations – phonon scattering.

Of these, 1 and 2 would seem to be, intuitively, the most likely at low temperatures under an applied electric field. Conditions for ternary

growth using molecular beam epitaxy (MBE) may promote 3 and 4 also, but the sources of 5 should be sufficiently suppressed.

It is remarkable that many of the above factors have a common derivation differing only in how the characteristic scattering potential is treated. With that in mind, a general derivation of the scattering rate will be given using the intermediate Fermi's Golden Rule[16] and the Born Approximation. The subsequent divergence when the characteristic scattering potentials are included will then follow.

2.4.2 Initial Considerations for Elastic Scattering

In this section, only the elastic scattering of electrons will be considered. This is the process whereby the momentum of the electron ($\hbar\mathbf{k}$) is generally changed, but not its energy (E). In an inelastic process, both these characteristics are generally changed.

As an electron or Bloch wave propagates in a real crystal, it is very likely that in the microscopic regime, this wave will be scattered into another momentum state by its interaction with a local perturbing potential. As stated in the numbered list in the introduction, these potentials have a wide range of possible sources and relative influence. With time, the probability of these scattered states existing increases linearly and hence, conversely, the probability of the existence of the pure Bloch wave eigenstates (ϕ_j) decreases. This linear increase is commonly expressed through the scattering rate. Eventually, one can visualise a complicated electron wave composed of many superposed states mixed in and remote from the Bloch 'ideal'.

Despite this, perturbation theory allows the use of the unperturbed Bloch wave states and its Hamiltonian ($\hat{H}_0(\underline{r},t)$). Owing to the linearity of Schrödinger's equation, the perturbing potential ($\hat{V}(\underline{r},t)$) can simply be added to form a modified Hamiltonian $[\hat{H}_0(\underline{r},t)+\hat{V}(\underline{r},t)]$ and hence yield the equation:

$$\left[\hat{H}_0(\underline{r},t)+\hat{V}(\underline{r},t)\right]\Psi(\underline{r},t)=i\hbar\frac{\partial}{\partial t}\Psi(\underline{r},t) \qquad (2.4.1)$$

where $\Psi(\underline{r},t)$ is the time-dependent wavefunction solution. Fermi's Golden Rule can be derived from equation 2.4.1, but as it is quite well known and often derived in textbooks,[3,16,17,18,19] I will merely state it here:

$$W_{fi}=\frac{2\pi}{\hbar}\sum_i|V_{fi}|^2\delta(\varepsilon_f-\varepsilon_i) \qquad (2.4.2)$$

where W_{fi} is the rate of change from the initial (i) to the final state (f); V_{fi} is the matrix representation of the perturbing potential which scatters an electron from its initial state with energy (ε_i) to its final state with energy (ε_f); δ is a delta function, which can be 'collision broadened'. The sum is over all initial states. However the matrix terms $|V_{fi}|^2$ should be approximately the same for all i as all the initial states are close in energy to the final states, as expressed by the delta-function. Therefore, the rate can be expressed as:

$$W_{fi}\cong\frac{2\pi|V_{fi}|^2}{\hbar}\sum_i\delta(\varepsilon_f-\varepsilon_i)=\frac{2\pi|V_{fi}|^2}{\hbar}N(\varepsilon_f) \qquad (2.4.3)$$

where $N(\varepsilon_f)$ is the density of states and ($\varepsilon_f=\varepsilon_i$).

2.4.3 Remote Ionised Impurity (RII) Scattering

Starting from the equation for Fermi's Golden Rule, equation 2.4.2/3, a reasonably accurate expression representing RII scattering can be derived. This form of scattering is found in modulation-doped layers, where significant direct ionised impurity scattering can be avoided. The great advantage conferred by the Golden Rule is its direct applicability not only to this form of scattering but to others, such as random alloy and interface-roughness. The treatment of the scattering potential (V_{fi}) is the major difference between these components.

V_{fi} is a general matrix element and has the form:

$$V_{fi} = \int \phi_f^* \hat{V} \phi_i d^2\underline{r} \qquad (2.4.4)$$

where the integral is over area; $\hat{V}$ is the scattering potential energy operator; ϕ_i and ϕ_f are the initial and final eigenstates, respectively, each of which can be defined as:

$$\phi_i = A^{-1/2} \exp(i\,\underline{k} \,.\, \underline{r}) \qquad (2.4.5)$$

$$\text{and } \phi_f = A^{-1/2} \exp(i\,(\underline{k}+\underline{q}).\underline{r}) \qquad (2.4.6)$$

where A is the area under consideration; $\boldsymbol{r}$ is the positional vector; $\boldsymbol{k}$ is the initial wavevector and $\boldsymbol{k}+\boldsymbol{q}$ is the final wavevector. Substituting 2.4.5 and 2.4.6 into 2.4.4, the following expression for V_{fi} is obtained:

$$V_{fi} = A^{-1} \int V(\underline{r}) \exp(-i\,\underline{q} \,.\, \underline{r})\, d^2\underline{r} \qquad (2.4.7)$$

where $\hat{V}$ is given its positional dependence ($V(\mathbf{r})$). Equation 2.4.7 shows that V_{fi} and $V(\underline{r})$ are linked through a 2D Fourier Transform ($\tilde{V}$), that is:

$$V_{fi} = A^{-1} \tilde{V}(\underline{q}) \qquad (2.4.8)$$

Substituting this potential into the expression for the Golden rule, equation 2.4.2 and emphasising that the initial and final states are represented by $\boldsymbol{k}$ and $\boldsymbol{k}+\boldsymbol{q}$, respectively, results in the following:

$$W_{\underline{k}+\underline{q},\underline{k}} = A^{-2}\frac{2\pi}{\hbar}\left|\tilde{V}(\underline{q})\right|^2 \delta(\varepsilon(\underline{k}+\underline{q})-\varepsilon(\underline{k})) \qquad (2.4.9)$$

This is the so-called 'Born Approximation' and expresses the proportionality of the scattering rate on the modulus squared of the Fourier transform of the scattering potential. Note also the A^{-2} proportionality, as this expression applies to one scattering centre and one final state. Assuming all final states are empty (ignoring Fermi factors), the total scattering rate ($1/\tau_q$) of a given centre will be the sum over $\boldsymbol{q}$ of all the individual rates ($W_{\mathbf{k+q,\,k}}$):

$$\frac{1}{\tau_q} = \sum_{\underline{q}} W_{\underline{k}+\underline{q},\underline{k}} \qquad (2.4.10)$$

The sum tends to the limit expressed by:

$$\frac{1}{\tau_q} \rightarrow \frac{A}{(2\pi)^2}\int W_{\underline{k}+\underline{q},\underline{k}}\, d^2\underline{q} \qquad (2.4.11)$$

Considering, now, all the scattering centres, whose number (N_{imp}^{2D}) is equal to the product of the area of the system and its centre density (n_{imp}^{2D}):

$$N_{imp}^{2D} = An_{imp}^{2D} \qquad (2.4.12)$$

Including these two additional factors, a more general expression emerges, where there is now no area dependence:

$$\frac{1}{\tau_q} = An_{imp}^{2D}\frac{A}{(2\pi)^2}\int \frac{1}{A^2}\frac{2\pi}{\hbar}\left|\tilde{V}(\underline{q})\right|^2 \delta(\varepsilon(\underline{k}+\underline{q})-\varepsilon(\underline{k}))d^2\underline{q} \qquad (2.4.13)$$

$$\Rightarrow \frac{1}{\tau_q} = \frac{2\pi n_{imp}^{2D}}{\hbar}\int\left|\tilde{V}(\underline{q})\right|^2 \delta(\varepsilon(\underline{k}+\underline{q})-\varepsilon(\underline{k}))\frac{d^2\underline{q}}{(2\pi)^2} \qquad (2.4.14)$$

In the derivation of equation 2.4.14, all scattering angles, including zero, were given equal weighting in the summing process. This leads to the 'quantum or single particle scattering time' or rate and hence mobility. Intuitively, it would be expected that scattering through large angles, that is, large ***q*** would affect the overall electron current somewhat more than that due to small angle scattering. To illustrate this, a general scattering diagram is shown below:

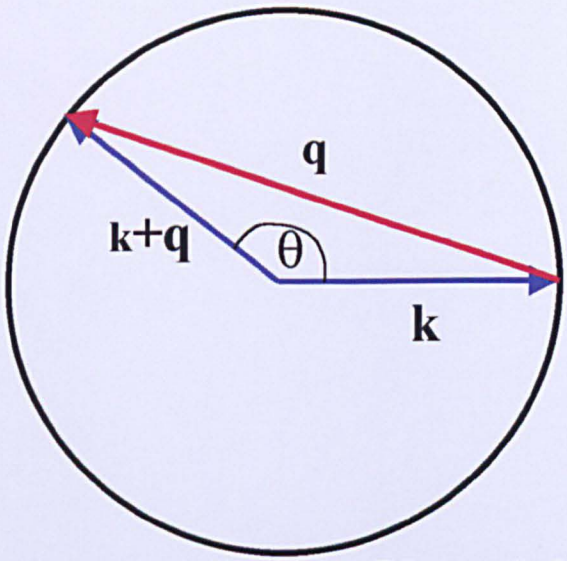


Figure 2.4.1 Scattering Diagram in **k**-space – **k** is the initial wavevector; **k+q** is the final wavevector and **q** is the resultant. *Sin(θ/2)=|**q**|/2|**k**|.*

A circle envelops all the state vectors, as energy must be conserved (|**k**|=|**k+q**|) and, therefore, it is easy to show that:

$$\left|\underline{q}(\theta)\right| = 2\left|\underline{k}\right|\sin(\theta/2) \qquad (2.4.15)$$

To allow for larger-angle weighting, an additional (1-cosθ) factor could be included, which will suppress smaller-angle scattering. Using the relation $cos^2(\theta/2) + sin^2(\theta/2) = 1$, (1-cosθ) can be re-expressed as:

$$1-\cos\theta = \frac{\left|\underline{q}\right|^2}{2\left|\underline{k}\right|^2} \qquad (2.4.16)$$

and equation 2.4.14 can be re-expressed on this occasion in terms of the 'transport scattering time' (τ_{tr}) as:

$$\frac{1}{\tau_{tr}} = \frac{2\pi n_{imp}^{2D}}{\hbar} \int \frac{|\underline{q}|^2}{2|\underline{k}|^2} \left|\tilde{V}(\underline{q})\right|^2 \delta(\varepsilon(\underline{k}+\underline{q}) - \varepsilon(\underline{k})) \frac{d^2\underline{q}}{(2\pi)^2} \qquad (2.4.17)$$

This scattering rate is related to the measured mobility and the $|q|^2$ term weights the expression towards large-angle scattering.

If $\tilde{V}(\underline{q})$ is independent of q (isotropic scattering), then $\tau_q = \tau_{tr}$, but if, as in the case of RII scattering, it is dependent on q, they will differ, often by large factors.

By changing the variables to that of the final state ($|\mathbf{k}'|$), where $k' = k+q$, and converting to polar co-ordinates, based on a k axis gives:

$$\frac{1}{\tau_{tr}} = \frac{n_{imp}^{2D}}{2\pi\hbar} \int_{-\pi}^{\pi} d\theta \int_0^{\infty} |\underline{k}'| d|\underline{k}'| \frac{|\underline{k}'-\underline{k}|^2}{2|\underline{k}|^2} \left|\tilde{V}(|\underline{k}'-\underline{k}|)\right|^2 \delta\left(\frac{\hbar^2|\underline{k}'|^2}{2m^*} - \frac{\hbar^2|\underline{k}|^2}{2m^*}\right) \qquad (2.4.18)$$

where the argument of the delta function is also in terms of $|k|$. As $|k| = |k|$, then $|k'-k| = 2|k| \sin(\theta/2)$ and the argument of the delta function appears to be zero. However, the influence of the delta function is retained by dividing equation 2.4.18 by the argument's derivative, that is $\hbar^2|k|/m^*$, to obtain:

$$\frac{1}{\tau_{tr}} = \frac{m^* n_{imp}^{2D}}{\pi\hbar^3} \int_0^{\pi} \left|\tilde{V}\left(2|\underline{k}|\sin(\theta/2)\right)\right|^2 (1-\cos\theta) d\theta \qquad (2.4.19)$$

where the integral with respect to $|k|$ has been evaluated. This can be re-expressed in terms of the wavevector q:

$$\frac{1}{\tau_{tr}} = \frac{m^* n_{imp}^{2D}}{2\pi\hbar^3|\underline{k}_F|^3} \int_0^{2|k_F|} \left|\tilde{V}(\underline{q})\right|^2 \frac{|\underline{q}|^2 d|\underline{q}|}{\sqrt{1-\left(|\underline{q}|/2|\underline{k}_F|\right)^2}} \qquad (2.4.20)$$

where $|k_F|$ replaces $|k|$ because of the degeneracy of the electron gas in these layers and $2|k_F|$ is, of course, the maximum magnitude q can have.

Dividing the integral by $|q|^2/2|k|^2$, that is, the factor (1-cos θ) will give back the quantum or single particle scattering rate ($1/\tau_q$).

If, however, the ionised donor centres are remote from the 2DEG, as is the case for delta-doped layers, then other factors should be included. The two principal factors are the potential smoothing effect of the intervening semiconductor, as expressed by the relative permittivity (ε_r) factor and the 2DEG electrons themselves. It can be shown that the potential, including such screening ($\tilde{V}_{scr}(\underline{q})$) has the following form:[3]

$$\tilde{V}_{scr}(q) = \frac{e^2}{2\varepsilon_0\varepsilon_r} \frac{e^{-|\underline{q}||d|}}{|\underline{q}| + |\underline{q}|_{TF}} \tag{2.4.21}$$

where q_{TF} is the Thomas-Fermi wavevector, which is equal to $2/a_B$, where a_B is the effective Bohr radius and d is the distance between the delta-layer and the 2DEG (the spacer width). Substituting this potential into equation 2.4.20 gives:

$$\frac{1}{\tau_{tr}} = \frac{m^* n_{imp}^{2D}}{2\pi\hbar^3 |\underline{k}_F|^3} \left(\frac{e^2}{2\varepsilon_0\varepsilon_r}\right)^2 \int_0^{2|k_F|} \frac{\exp(-2|\underline{q}||d|)}{\left(|\underline{q}| + |\underline{q}|_{TF}\right)^2} \frac{|\underline{q}|^2 d|\underline{q}|}{\sqrt{1 - \left(|\underline{q}|/2|\underline{k}_F|\right)^2}} \tag{2.4.22}$$

An even better equation can be obtained, by accounting for the distribution over depth of the electron wavefunction, which is not uniform with depth. Electrons at different depths will experience different potentials. To achieve this, the potential in matrix form (V_{nm}) should be calculated using wavefunctions distributed over depth. The Fang-Howard variational wavefunction (u(z)) provides a reasonable approximate wavefunction for the first subband:

$$u(z) = \left(\frac{1}{2}b^3\right)^{1/2} z \exp\left(-\frac{1}{2}bz\right) \tag{2.4.23}$$

where *b* is a parameter and *z* represents depth measured from the channel-spacer boundary towards the substrate. This wavefunction makes a reasonable approximation as the channel energy well approximates a triangle at lower energies – see fig.1.2.4.

The scattering potential matrix term (V_{nm}), where *n* and *m* represent final and initial subband states, can be calculated from the following:

$$V_{nm} = \frac{1}{A}\int u_n^*(z)\exp(-i(\underline{k}+\underline{q}).\underline{r})V(\underline{r},z)u_m(z)\exp(i(\underline{k}.\underline{r})\,d^3\underline{r} \quad (2.4.24)$$

$$= \frac{1}{A}\int u_n^*(z)u_m(z)dz\int V(\underline{r},z)\exp(-i\underline{q}.\underline{r})\,d^2\underline{r} \quad (2.4.25)$$

$$= \frac{1}{A}\int u_n^*(z)\,u_m(z)\,\tilde{V}(|\underline{q}|,z)\,dz \quad (2.4.26)$$

where $\tilde{V}$ is the Fourier transform. A comparison between V_{nm} and $\tilde{V}$ at $z = 0$, through the expression:

$$V_{nm} = \frac{1}{A}F_{nm}(|\underline{q}|)\,\tilde{V}(|\underline{q}_0|) \quad (2.4.27)$$

where $F_{nm}(q)$ is a so-called 'form factor', allows the effect of the wavefunction to be evaluated. When intra-subband scattering (n = m), and $q = 0$ applies, then $F_{nm} = 1$. Substitution of the Fang-Howard wavefunction into equation 2.4.24 implies that:

$$F_{11}(|\underline{q}|) = \left(\frac{b}{b+|\underline{q}|}\right)^3 \quad (2.4.28)$$

Likewise, the Thomas-Fermi wavevector ($|\mathbf{q}|_{TF}$)in equation 2.4.21, will also be required to be modified by a form factor. (A thin or delta distribution of electrons was also assumed in its original derivation.) This revised form factor ($G(\mathbf{q})$) is given by:

$$G(|\underline{q}|) = \frac{1}{8}\left[2\left(\frac{b}{b+|\underline{q}|}\right)^3 + 3\left(\frac{b}{b+|\underline{q}|}\right)^2 + 3\left(\frac{b}{b+|\underline{q}|}\right)\right] \tag{2.4.29}$$

and the final rate equation for RII scattering becomes:

$$\frac{1}{\tau_{tr}} = \frac{mn_{imp}^{2D}}{2\pi\hbar^3|\underline{k}_F|^3}\left(\frac{e^2}{2\varepsilon_0\varepsilon_r}\right)^2 \int_0^{2|\underline{k}_F|} \frac{\exp(-2|\underline{q}||d|)}{\left(|\underline{q}| + |\underline{q}|_{TF}\, G(|\underline{q}|)\right)^2}\left(\frac{b}{b+|\underline{q}|}\right)^6 \frac{|\underline{q}|^2\, d|\underline{q}|}{\sqrt{1-\left(|\underline{q}|/2|\underline{k}_F|\right)^2}} \tag{2.4.30}$$

where $G(|\underline{q}|)$ is given in equation 2.4.29.

2.4.4 Random Alloy Scattering

2.4.4.1 Introduction

Often binary compounds do not offer enough flexibility in, for example, band-gap engineering. As discussed earlier, inter-mixing a third or even a fourth element during growth can access intermediate characteristics, although the degree of mixing is limited somewhat to avoid lattice relaxation. Such incorporation, however, will promote crystalline disorder, as the different atoms belonging to the same group are randomly distributed on their sub-lattice. Therefore, although the physical nature of the crystal remains periodic, this disorder gives rise to a non-periodic potential, which the electrons will experience. An approximate analytical expression for the scattering rate can be formulated for the periodic part, although the modelling of the random alloy scattering potential has been and remains contentious.

Unlike RII scattering in modulation doped systems, this form of scattering is much more direct, as its origin is intrinsic to the channel. In a real system, the random distribution of these atoms is unlikely; the distribution will depend on the growth temperature and the relative atomic proportions. Higher growth temperature tends to segregate one of the atoms, for example indium, into clusters, which can 'float' upwards with continued growth. The pseudomorphic strain may also advance this effect and the possibility of transport anisotropy emerges if the crystalline morphology and/or strain influence the spatial distribution of these clusters.

Clustering will, of course, reduce the applicability of an analytical expression derived on the assumption that the atoms are randomly distributed, but a comparison between this data and that obtained experimentally can prove useful. A comparatively large difference would indicate that heavy clustering is present. A derivation expressing mobility due to **random** alloy scattering will now follow.

2.4.4.2 A Derivation Representing Random Alloy Disorder Scattering

As the scattering potential is non-periodic, the Bloch Theorem and derivations therefrom are not applicable. However, this potential, (H) can be expressed as the linear sum of the periodic potential (H_0) and that due to the random fluctuation (H'). This was the same concept applied in the derivation of the RII scattering time, equation 2.4.30, via Fermi's Golden rule, equations 2.4.2/3. The matrix element (V_{fi}) in the rule, incorporating the scattering potential (ΔU), is given by:

$$V_{fi} = \int \exp(\underline{k}_f - \underline{k}_i) \cdot \underline{r}\, \Delta U(\underline{r})\, d^3\underline{r} \qquad (2.4.31)$$

The main difference between this scattering component and the previous RII type, is the characterisation of this potential. H_0 can be regarded as the average potential of the crystal alloy and H' the difference between this potential and the actual local potential.

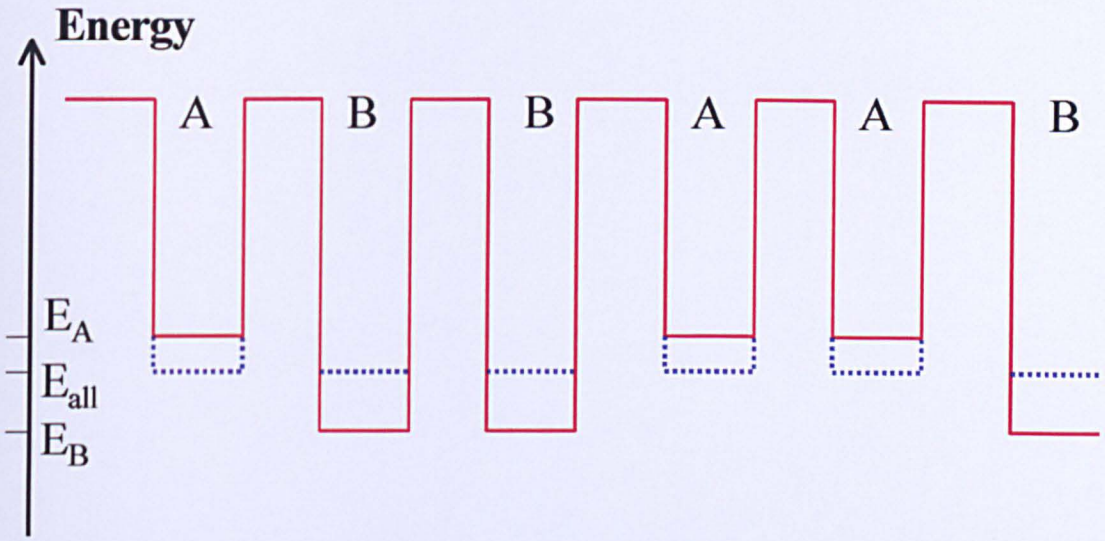


Figure 2.4.2 Schematic energy diagram showing, comparatively, the various scattering potentials of atoms *A*, and *B* and the weighted average of the two, taking relative abundance into account (E_A,E_B and E_{all} respectively). (after Singh[17])

The 'Virtual Crystal Approximation' can be used to express the average potential to the first order. In this, the average periodic potential (E_{all}) is a weighted sum of the local potentials E_A and E_B due to the alloy atoms *A* and *B*, according to their relative abundance, *x* and *1-x*, respectively:

$$E_{all} = xE_A + (1-x)E_B \qquad (2.4.32)$$

The actual varying potential of the crystal can be represented schematically in the following way in fig. 2.4.2. Note the comparative influence of either group III atom, *A* or *B*, on the actual potential (solid

line) and its difference from the average periodic virtual crystal potential (dotted line). Depending on the relative abundance of *A* to *B*, then from equation 2.4.32 above, the dotted line will lie nearer the actual potential of the more abundant species.

Therefore, it is now assumed that this difference between the actual and the virtual crystal potential represents the highly localised scattering potential within a unit cell. For species *A*, this energy difference (E_{all} - E_A) or local perturbing potential (H′ or U_0) can be expressed as:

$$E_{all} - E_A = xE_A + (1-x)E_B - E_A \tag{2.4.33}$$

$$\Rightarrow E_{all} - E_A = (1-x)[E_B - E_A] \tag{2.4.34}$$

Similarly, for atom *B*:

$$E_{all} - E_B = x[E_B - E_A] \tag{2.4.35}$$

As implied in the previous figure (2.4.2), the potential ((E_{all}-$E_{A/B}$) or U_0) is assumed to be constant over the size of a unit cell (r_0) with an abrupt change at the edge:

$$\begin{aligned} \Delta U(\underline{r}) &= U_0 \qquad |\underline{r}| \le r_0 \\ &= 0 \qquad |\underline{r}| > r_0 \end{aligned} \tag{2.4.36}$$

Invoking the Born Approximation, the rate (W(**k**)) becomes:

$$W(\underline{k}) = \frac{2\pi}{\hbar} \sum_{k'} |M_{kk'}|^2 \, \delta(\varepsilon_k - \varepsilon_{k'}) \tag{2.4.37}$$

with:

$$M_{kk'} = \int \exp[i(\underline{k} - \underline{k}').\underline{r}] \, \Delta U(\underline{r}) \, d^3\underline{r} \tag{2.4.38}$$

The approximation:

$$\exp[i(\underline{k} - \underline{k}').\underline{r} \approx 1 \tag{2.4.39}$$

can be introduced as the scattering potential is highly localised (over a unit cell) compared with $1/k_F$. Therefore:

$$M_{\underline{k}\,\underline{k}'} = \frac{4\pi}{3} r_0^3 U_0 \tag{2.4.40}$$

The rate now becomes:

$$W(\underline{k}) = \frac{2\pi}{\hbar}\left(\frac{4\pi}{3} r_0^3 U_0\right)^2 \frac{1}{(2\pi)^3} \int \delta(\varepsilon_{\underline{k}} - \varepsilon_{\underline{k}'})\, d^3\underline{k}' \tag{2.4.41}$$

$$= \frac{2\pi}{\hbar}\left(\frac{4\pi}{3} r_0^3 U_0\right)^2 N(\varepsilon_{\underline{k}}) \tag{2.4.42}$$

where $N(\varepsilon_k)$ is the density of states.

For a face-centred cube lattice, $r_0 = a\sqrt{3}/4$, where a is the cubic edge length. The unit (cell) volume (V_0) is $a^3/4$. The relation between r_0 and V_0 is given by:

$$\left(\frac{4\pi}{3} r_0^3\right)^2 = \frac{3\pi^2}{16} V_0^2 \tag{2.4.43}$$

The rate now emerges as:

$$W(\underline{k}) = \frac{2\pi}{\hbar}\left(\frac{3\pi^2}{16} V_0^2\right) U_0^2\, N(\varepsilon_{\underline{k}}) \tag{2.4.44}$$

From equations 2.4.35/36, $U_0 = x(E_A - E_B)$ or $(1-x)(E_A - E_B)$ depending on which atomic species is being considered. Renaming (E_A-E_B) as U_{all}, then the scattering rate for the atom (A) with proportion x is:

$$W_A(\underline{k}) = \frac{2\pi}{\hbar}\left(\frac{3\pi^2}{16} V_0^2\right) x^2 U_{all}^2\, N(\varepsilon_{\underline{k}}) \tag{2.4.45}$$

with a similar expression for atom B, but with the proportion (1-x) replacing x. There are x/V_0 of type A and $(1-x)/V_0$ of type B, and if the scattering rates of the two species can be added, then the rate becomes:

$$W_{Total} = \frac{2\pi}{\hbar}\left(\frac{3\pi^2}{16} V_0\right) U_{all}^2\, N(\varepsilon_{\underline{k}})\left[x(1-x)^2 + (1-x)x^2\right] \tag{2.4.46}$$

$$\Rightarrow W_{Total} = \frac{3\pi^3}{8\hbar} V_0\, U_{all}^2\, x(1-x)\, N(\varepsilon_{\underline{k}}) \tag{2.4.47}$$

2.5 Localised Transport

2.5.1 Introduction

The characteristics of electron transport in a 2DEG at high resistance has occupied much theoretical and experimental research over the years. Mott conceived the idea of a minimum metallic conductivity, when the resistance approached about 10kΩ per square.[20] This concept suggests the idea of an abrupt metal-insulator transition, but more recent studies[21] have rather supported the presence of a gradual transition (at around 10kΩ per square) from the strongly exponential localisation at higher resistance to the weakly logarithmic localisation in the so-called 'metallic' regime. In other words, localisation always exists (no extended states), with only the degree changing.

Temperature measurements supported this transitional behaviour,[22,23,24] although the effect of correlations could not be ruled out.[25] This latter factor, however, can be isolated from localisation by making use of both parallel and perpendicular magnetic fields.[26,27] It is difficult to measure localisation at high conductivity in the metallic regime, and therefore the conductivity regions either side of, but close to, the 10kΩ per square mark, receive the most attention, as indeed they do in this work. From the discussion above, it is useful to define the two regions as the weak and strong localisation regimes, each of which will be considered, in turn.

2.5.2 Weak Localisation

2.5.2.1 General Description and Explanation

Experimental observations under certain conditions of low magnetic field, temperature and carrier density, suggested that the semi-classical Drude conductivity under an applied electric field was no longer applicable. Under these conditions, a lower conductivity or a higher resistivity than that predicted by Drude was observed. To account for this, a quantum explanation was developed, although the carrier trajectories can be treated classically. (See, for example, van Houten et al[28] (1986) for the trend at low temperature.)

Ignoring presently the influence of a magnetic field, the temperature dependence of the resistivity can be explained as follows. As the temperature of a 2DEG falls, the ratio of inelastic to elastic collisions, involving carrier electrons also falls. At very low temperatures there is virtually no inelastic scattering and the residual mobility/ conductivity is thus dominated by elastic scattering factors, see section 2.4. Impurities and/or crystal defects usually form the major scattering centres. Therefore classically, a monotonic fall in resistance with temperature would be expected. However, as the above conditions are approached, a small rise is in fact observed and its magnitude is emphasised when the 2DEG is further confined to narrow channels.

In the semi-classical Drude derivation, the idea of a velocity correlation function is used, which arises from the elastic scattering events. This function decays exponentially with the mean scattering time (τ). However, if it is noted that purely elastic collisions do not destroy

the phase of the electrons' wavefunction, then a more convincing quantum mechanical argument can be introduced for transport under these special conditions. It should be emphasised that the Drude construction is adequate under most conditions, but that a quantum-mechanical correction may be required on occasion.

Modelling electrons as waves allows the concept of wave interference arising from long-range correlations. A noteable effect of this quantum interference is the enhanced probability for back-scattering of the electron-waves. This phenomenon was named 'weak localisation' to distinguish it from the stronger variant observed at higher resistance,[21,29] with the quantum correction named as the 'weak localisation correction'.

A number of theoretical models for weak localisation have been presented, including those using diagrammatic methods[29] and others using perturbation theory.[30] However, an easier approach was found when the concept of coherent back-scattering,[31,32] combined with path integrals[33] ("Feynmann Paths"), was advanced. Back-scattered electrons can be represented schematically as in fig. 2.5.1. The initial and final positions are represented by $\boldsymbol{r}$ and $\boldsymbol{r'}$, respectively, and the classical trajectory shown is just one of the number i possible in a time t. Classical trajectories are allowable if the electrons' Fermi wavelength is much less than the mean scattering length (l). As can be seen in the figure, the electron has returned to $\mathbf{r}$ ($\mathbf{r'}$) after encountering 7 elastic collisions. The total probability of back scattering is given by:[34]

$$P(r,r',t)=\left|\sum_i A_i\right|^2=\sum_i |A_i|^2+\sum_{i\neq j} A_i A_j \qquad (2.5.1)$$

where A_i is the probability amplitude of the i^{th} trajectory.

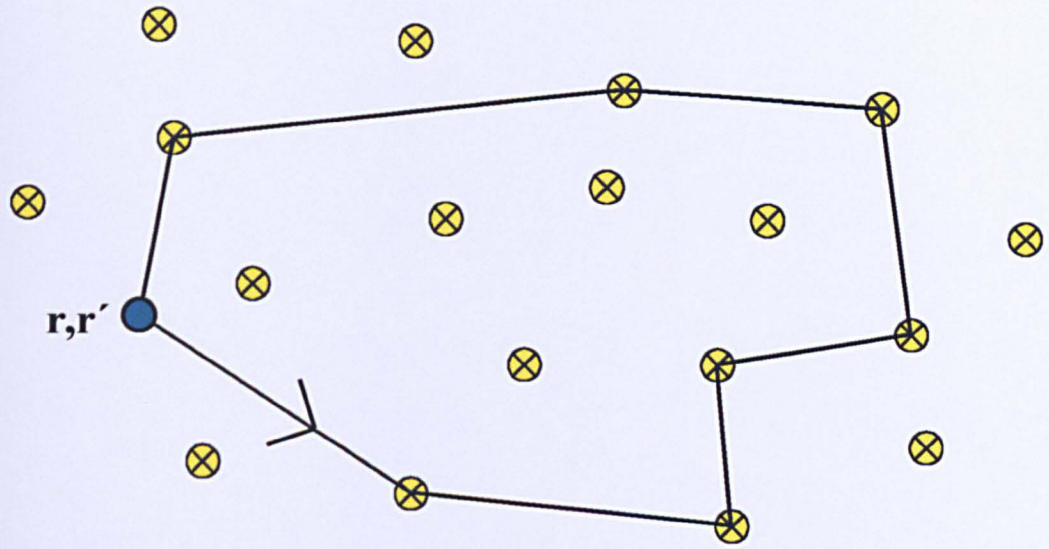


Figure 2.5.1 A possible back-scattered classical trajectory of an electron, which has encountered 7 elastic scattering events.

Quantum interference is represented in the second term of the extreme right-hand-side (RHS), whilst the first term of the RHS accounts for classical diffusion. If $\boldsymbol{r}$ and $\boldsymbol{r}'$ are not coincident then, generally, the electron waves are uncorrelated. In a large enough sample, there are many possible trajectory paths and hence times between the two parts and the interference term averages to zero. Sample-dependent conductivity can often be observed if this interference term does not average out, either if the sample is small (mesoscopic) or under special circumstances found much more generally in samples of all sizes.

One special circumstance arises when the reverse trajectory, $\boldsymbol{r}'$ to $\boldsymbol{r}$ is considered. This path is equally likely as the one shown above, with the same time-of-flight (t). Time reversal invariance ensures that the probability amplitude of each of the electron waves of the two counter-propagating paths is equal. It turns out that the understanding benefits from a consideration of these counter-propagating pairs. Applying equation 2.5.1 to the equal amplitudes A^+ and A^- representing these

trajectories gives double the back-scattering probability when compared with a treatment using only the classical diffusion term. An equal time of flight implies that the time-dependent phase difference ($Et/\hbar$), where E is the energy, is identical for both paths. Each collision preserves the magnitude of $\boldsymbol{k}$ ($|\mathbf{k}|$), hence preserving E. Also the spatial phase difference ($\Sigma \boldsymbol{k}.\boldsymbol{r}$ around the closed trajectory) is the same because $|\boldsymbol{k}|$ is preserved and the paths are identical.

Therefore, if the above conditions, in addition to those listed at the beginning of this sub-section, are met, an enhanced back-scattering of electron waves can lower the conductivity due to an applied electric field. Thus the classical diffusion constant is reduced by the inherent disorder within the system. Methods of varying this weak localisation prove useful in its study, and two of these will now be considered.

2.5.2.2 The Destruction of Weak Localisation

There are two principal methods used to vary the conditions for weak localisation. These are the application of a magnetic field, and temperature, and both will be considered in turn.

The principal relevant effect of a magnetic field is to destroy time reversal invariance. With the directional Lorentz force acting on electrons, it is unlikely that counter-propagating path pairs will have the same time-of-flight. Indeed, a phase difference (ϕ), the Aharonov-Bohm phase difference, develops between each member of the pair. Not surprisingly, this is magnetic field dependent and is given by:[34]

$$\phi = \frac{2eBS}{\hbar} \equiv \frac{2S}{l_m^2} \equiv 4\pi \frac{\Phi}{\Phi_0} \qquad (2.5.2)$$

where S is the enclosed area of the trajectories; l_m is the magnetic length $\{(\hbar/eB)^{1/2}\}$; Φ is the enclosed flux and Φ_0 is the flux quantum (h/e). Clearly, if S or B gets too large, then the phase difference developed will destroy the enhanced back-scattering and hence weak localisation. At certain defined characteristic magnetic fields, the phase difference appropriate to certain relaxation mechanisms will be such as to "destroy" the phase coherence of the counter-propagating trajectories. These characteristic fields are the transport, phase and spin-orbit relaxation fields (B_{tr}, B_ϕ and B_{so}, respectively). They are related to the corresponding transport, phase and spin-orbit relaxation times (τ, τ_ϕ, and τ_{so}, respectively), and are quoted in Zduniak et al[35] (1995), which also discusses weak localisation in layers with an InGaAs channel. The characteristic fields are:

$$B_{tr} = \frac{\hbar}{4eD\tau}; \quad B_\phi = \frac{\hbar}{4eD\tau_\phi}; \quad B_{so} = \frac{\hbar}{4eD\tau_s} \tag{2.5.3}$$

where D is the classical diffusion constant and is equal to $l^2/2\tau$, where l is the elastic mean free path. In deriving the weak localisation correction ($\Delta\sigma(B)$), this paper quotes the appropriate expression given by Dyakonov[36] (1994), which ignores the spin orbit and phase relaxation:

$$\Delta\sigma(B) = -\frac{e^2}{2\pi^2\hbar}F(x) \tag{2.5.4}$$

where

$$F(x) = x\sum_{n=0}^{\infty}\frac{P_n^3}{1-P_n} \tag{2.5.5}$$

with $x=B/B_{tr}$ and

$$P_n = s\int_0^{\infty} dt\, \exp(-st - t^2/2)L_n(t^2) \tag{2.5.6}$$

with $s=(2/x)^{1/2}$ and L_n the n^{th} Lageurre polynomial.

Phase relaxation can be accounted for by the incorporation of an additional exponential factor in P_n, which then becomes:

$$P_n = s\int_0^\infty dt \exp[-s(1+\beta_\phi)t - t^2/2]L_n(t^2) \qquad (2.5.7)$$

where $\beta_\phi = \tau/\tau_\phi = B_\phi/B_{tr}$ and the function $F(x)$ now becomes $F(x,\beta_\phi)$.

Spin relaxation can be introduced through a generalisation of the result quoted by Ilikama et al,[37] (1980). In a 2DEG subjected to a fluctuating magnetic field, the spin component of the electrons perpendicular to this plane, relax at twice the rate of the two in-plane components. Their time constant is given by τ_s. Including this process, equation 2.5.4 extends to:

$$\Delta\sigma(B) = -\frac{e^2}{2\pi^2\hbar}\left[F(x,\beta_{s1}) + \frac{1}{2}F(x,\beta_{s2}) - \frac{1}{2}F(x,\beta_\phi)\right] \qquad (2.5.8)$$

where

$$\beta_{s1} = \frac{B_\phi + B_{so}}{B_{tr}};\quad \beta_{s2} = \frac{\beta_\phi + 2B_{so}}{B_{tr}};\quad \beta_\phi = \frac{B_\phi}{B_{tr}} \qquad (2.5.9)$$

At small fields ($x=B/B_{tr}<<1$), equation 2.5.8 reduces to:

$$\Delta\sigma(B) = -\frac{e^2}{2\pi^2\hbar}\left[\Psi\left(\frac{1}{2}+\frac{1}{x}\right) - \Psi\left(\frac{1}{2}+\frac{\beta_{s1}}{x}\right) + \frac{1}{2}\Psi\left(\frac{1}{2}+\frac{\beta_\phi}{x}\right) - \frac{1}{2}\Psi\left(\frac{1}{2}+\frac{\beta_{s2}}{x}\right)\right] \qquad (2.5.10)$$

where Ψ is the di-gamma function. At large fields (x>>1), equation 2.5.4 is recovered, with the magneto-conductance described as "universal...for a given elastic scattering mechanism". In other words, without the low-field complications of spin and phase relaxation, the trend in magneto-conductance is uniform, regardless of the variation in the other sample parameters. Below $B=B_{tr}$, the conductivity trend in various samples is non-uniform, and this is shown both by experiment and theory in the paper.

The other main way of affecting weak localisation is through the application of heat energy. In demonstrating this phenomenon, it is useful to start with the equation for conductivity (σ) derived by Abrahams et al[21] (1979). It is given by:

$$\sigma = \sigma_0 - A\frac{e^2}{\hbar}\ln\left(\frac{l_\phi}{l}\right) \qquad (2.5.11)$$

where A is a constant, and σ_0 is the Drude conductivity and l_ϕ is the phase coherence length. Bergmann[31,38] (1983) derived a similar equation using a circle in $\boldsymbol{k}$-space, which maps the evolution of the $\boldsymbol{k}$ vectors with each elastic scattering event. The equation for weak localisation, given in terms of the change in conductivity ($\Delta\sigma$), is:

$$\Delta\sigma = -\frac{e^2}{2\pi^2\hbar}\ln\left(\frac{\tau_\phi}{\tau}\right) \qquad (2.5.12)$$

The phase relaxation time (τ_ϕ) is, of course, intimately connected with the temperature (T) and this relation is expressed through the relation:

$$\frac{1}{\tau_\phi} \propto T^P \qquad (2.5.13)$$

where P is a constant equal to 1 for disordered 2D systems[39]. Using this relation in equation 2.5.12 gives:

$$\Delta\sigma(T) = +\frac{e^2}{2\pi^2\hbar}P\ln T \qquad (2.5.14)$$

Again, here, $\Delta\sigma(T)$ is the change in conductivity representing the effect of weak localisation. With increasing temperature, for a given carrier density, there is a gradual reduction in τ_ϕ due to the increase in the inelastic scattering rate. This factor reduces the likelihood of coherent back-scattering, and hence, weak localisation, but by implication, increases the conductivity. This equation implies that, where weak

localisation exists, a linear trend in conductivity with the natural log of the temperature should be apparent.

2.5.3 Strong Localisation

If a 2DEG, or other carrier population, is strongly depleted, the Fermi level also drops in energy. Here it may encounter a scalar potential field or other artifacts, which can render localisation, or even confinement, of the carriers as the norm. As stated in section 2.5.1, there is no abrupt change from the characteristics of weak localisation to those of strong localisation, but theoretical and experimental signatures of the transition do exist.

It is likely that in real samples, local or remote effects will perturb the conduction band-edge in the channel. Local effects include concentration fluctuations of certain atomic species in ternary compounds. Two examples of the remote effects include the impact of the remote ionised donors[40] at the level of the channel, and the charges thought to be found near the Si-SiO_2 interface in MOSFETs.[41] In either case, it is probable that in high channel depletion, when the Fermi level is low, a population of highly localised electrons would develop, see fig. 2.5.2. According to the Ioffe-Regel Criterion, this will occur when $|\boldsymbol{k}_F| < 1/2\pi l$, where l is the mean free path and $\boldsymbol{k}_F$ is the Fermi wavevector, which is proportional to the square root of the carrier density. As is shown in fig. 2.5.3, the density of states, when distributed over energy, will therefore, have a 'tail' instead of the sharp transition to the full 2D

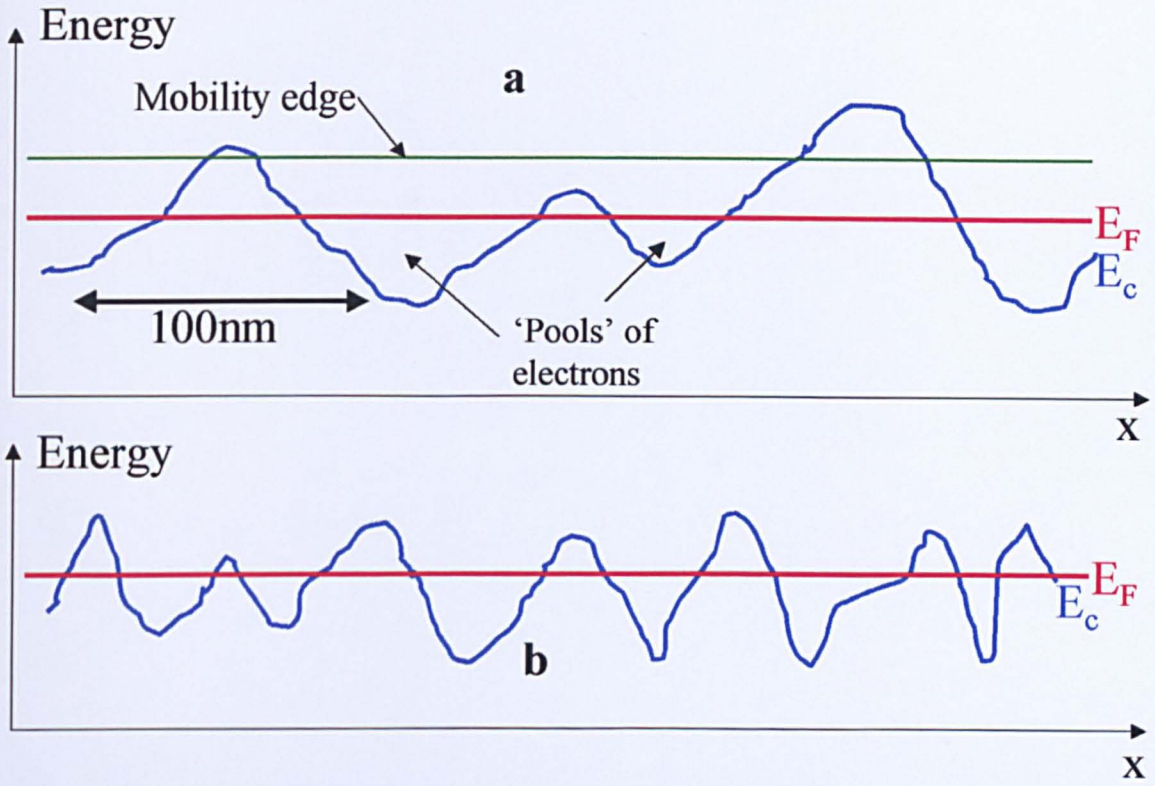


Figure 2.5.2. Schematic representation of potential energy fluctuations. **a**). Long-range fluctuations with 'pools' of localised electrons below the Fermi level. **b**). Short-range fluctuations. E_F is the Fermi level, and E_c the conduction band-edge. The mobility-edge will be discussed later.

density of states at E_c. These contain the localised states. This density is given by the expression $2\pi g_s g_v m^*/h^2$, where g_s is the spin degeneracy and g_v is the valley degeneracy. Also shown in this, and the previous figure is the so-called mobility-edge, which is the energy level regarded as separating the localised states of the 'hopping' regime from the essentially free, extended states of the metallic regime.

The energy required for an electron at the Fermi level to attain the mobility edge is called the activation energy (E_A). This activation energy, which is also dependent on carrier density, can be calculated from basic Arrhenius plots - a schematic part of which is shown in fig. 2.5.4. E_A can be calculated from the tangent gradients of these plots, as

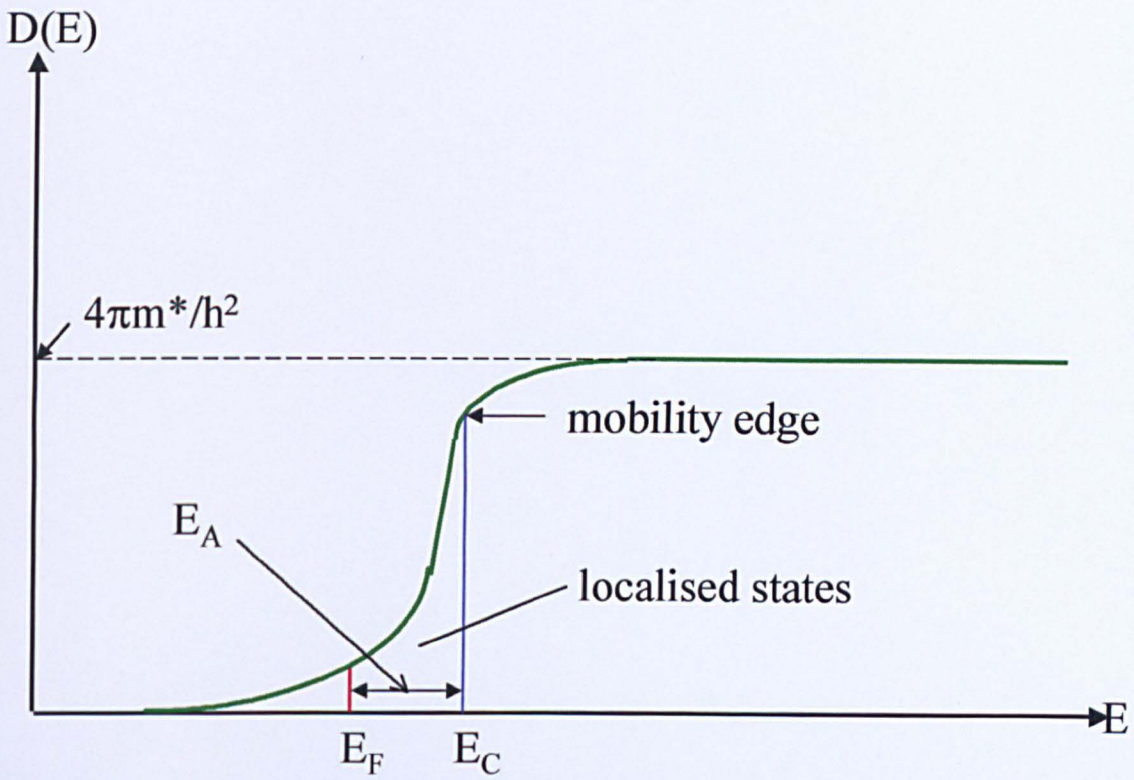


Figure 2.5.3. The 2D density of states distributed over energy, showing a tail containing the localised states found in real samples. The plateau represents the full metallic 2D density of states.

the conductivity (σ) is related to the activation energy in the following way:

$$\sigma = \sigma_0 \exp\left(\frac{-E_A}{k_B T}\right) \qquad (2.5.15)$$

where σ_0 is the minimum metallic conductivity pre-factor. The three lines of a given colour represent three different depletions or carrier densities at a given range of temperatures. The largest gradient represents activation from the lowest Fermi level of the three, and hence, the strongest localisation. The lines of red colour represent 'ideal' behaviour, in that they all extrapolate to one value of minimum metallic conductivity, (σ_0 or G_0), at infinite temperature. This phenomenon is believed to be due to short-range fluctuations attributed to charges near the interface, as described above. The minimum metallic conductivity

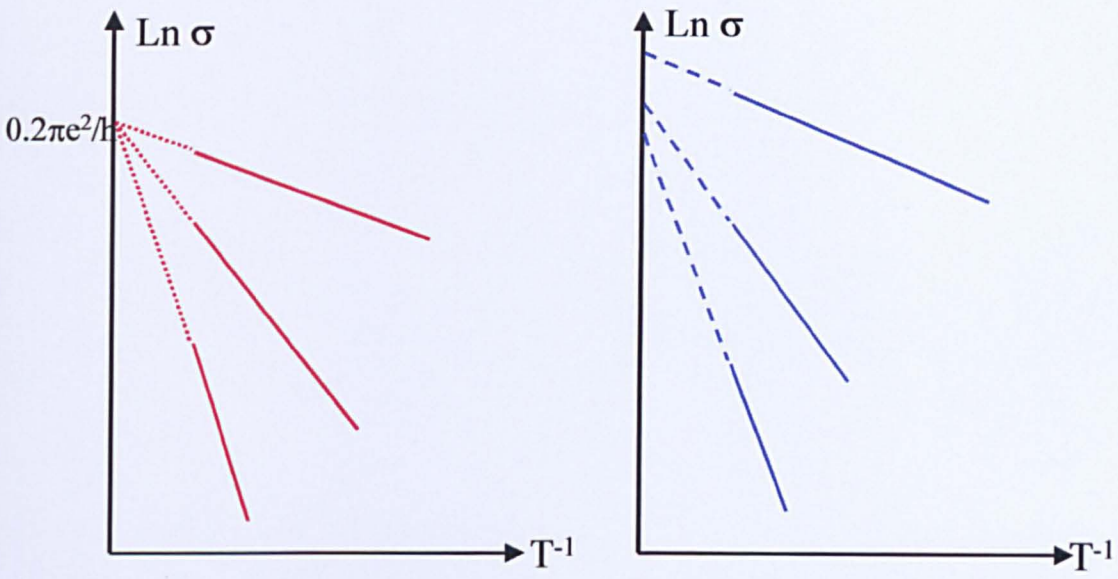


Figure 2.5.4. A schematic part-Arrhenius plot indicating 'ideal' behaviour - red lines, and 'non-ideal' behaviour - blue lines (see text).

is signified and is about $0.2\pi e^2/h$ and represents the mobility edge. However, 'non-ideal' behaviour, (the blue lines in fig. 2.5.4), is also likely when the perturbation predominantly involves long-range fluctuations. In this case, no single mobility-edge and a dependence on channel depletion, is apparent.

Percolation theories[42] have been partially successful in explaining 'non-ideal' behaviour, but Coulomb charging effects within, as well as tunnelling between, the depleting electron pools, must also play a role in the transport at low temperature. Eventually, contraction to single-state electron 'pools' at very high depletion should be possible, and hence, variable range hopping, with a $\log\sigma \propto T^{-1/3}$ characteristic in 2D systems, could be observed.

2.6 References

1. Long A.R., Davies J.H., Kinsler M., Vallis S., Holland M.C., Semiconductor Science and Technology, **8**, 1581, (1993).
2. Skuras E., Holland M.C., Barton C.J., Davies J.H., Long A.R., Semiconductor Science and Technology, **10**, 922, (1995).
3. Davies J.H., The Physics of Low-Dimensional Semiconductors, Cambridge University Press, (1998).
4. Stern F., Physical Review B, **5**, 4891, (1972).
5. Ando T., Fowler A.B., Stern F., Review of Modern Physics, **54**, 437, (1982).
6. Coleridge P.T., Stoner R., Fletcher R., Physical Review B, **39**, 1120, (1989).
7. Diaz-Paniagua C., Hidalgo M.A., Urbina A., Batallan F., Gilperez J.M., Fernandez de Avila S., Gonzalez-Sanz F., Muñoz E., Solid State Communications, **109**, 57, (1999).
8. van der Burgt M., Karavolas V.C., Peeters F.M., Singleton J., Nicholas R.J., Herlach F., Harris J.J., Van Hove M., Borghs G., Physical Review B, **52**, 12218, (1995).
9. Zrenner A., Koch F., Williams R.L., Stradling R.A., Ploog K., Weimann G., Semiconductor Science and Technology, **3**, 1203, (1988).
10. Sobkowicz P., Wilamowski Z., Kossut J., Semiconductor Science and Technology, **7**, 1155, (1992).
11. Buks E., Heiblum M., Levinson Y., Shtrikman H., Semiconductor Science and Technology, **9**, 2031, (1994).
12. Coleridge P.T., Semiconductor Science and Technology, **12**, 22, (1997).

13. Kastalsky A., Hwang J.C.M., Solid State Communications, **5**, 317, (1984).

14. Dmowski L.H., Zduniak A., Litwin-Staszewska E., Contreras S., Knap W., Robert J.L., Physica Status Solidi B, **198**, 283, (1996).

15. Koenraad P.M., Shi J.M., van de Stadt A.F.W., Smets A., Devreese J.T., Wolter J.H., Peeters F.M., Perenboom J.A.A.J., Superlattices and Microstructures, **21**, 237, (1997).

16. Fermi E., Nuclear Physics, University of Chicago Press, (1950).

17. Singh J., Quantum Mechanics, Fundamentals and Applications to Technology, Wiley, (1997).

18. Ridley B.K., Quantum Processes in Semiconductors, 3rd Edition, Clarendon Press, (1993).

19. Liboff R.L., Introductory Quantum Mechanics, 2nd Edition, Addison-Wesley (1992).

20. Mott N.F., Metal-Insulator Transition, Taylor and Francis, (1974).

21. Abrahams E., Anderson P.W., Licciardello D.C., Ramakrishnan T.V., Physical Review Letters, **42**, 673, (1979).

22. Dolan G.J., Osheroff D.D., Physical Review Letters, **43**, 72, (1978).

23. Bishop D.J., Tsui D.C., Dynes R.C., Physical Review Letters, **44**, 1153, (1980).

24. Giordano N., Physical Review B, **22**, 5635, (1980).

25. Al'tshuler B.L., Khmel'nitskii D., Larkin A.J., Lee P.A., Physical Review B, **22**, 5142, (1980).

26. Uren M.J., Davies R.A., Pepper M., Journal of Physics C, **13**, L985, (1980).

27. Bishop D.J., Dynes R.C., Tsui D.C., Physical Review B, **26**, 773, (1982).

28.van Houten H., van Wees B.J., Heijman M.G.J., Andre J.P., Applied Physics Letters, **49**, 1781, (1986).

29. Lee P.A., Ramakrishnan T.V., Review of Modern Physics, **57**, 287, (1985).

30. Anderson P.W., Abrahams E., Ramakrishnan T.V., Physical Review Letters, **43**, 718, (1979).

31. Bergmann G., Physical Review B, **28**, 2914, (1983).

32. Khmel'nitskii D., Physica, **126B**, 235, (1984).

33. Chakravarty S., Schmid A., Physics Reports, **140**, 193, (1986).

34. Beenakker C.W.J., van Houten H., Solid State Physics, **44**, 1, (1991).

35. Zduniak A., Dyakonov M.I., Litwin-Staszewska E., Knap W., Acta Physica Polonica A, **88**, 1053, (1995).

36. Dyakonov M.I., Solid State Communications, **92**, 711, (1994).

37. Ilikama S., Larkin A.I., Nagaoka Y., Progress in Theoretical Physics, **63**, 781, (1980).

38. Bergmann G., Physics Reports, **107**, 1, (1984).

39. Emeleus C.J., Ph.D. Thesis, University of Warwick, (1993).

40. Nixon J.A., Davies J.H., Physical Review B, **41**, 7929, (1990).

41. Ando T., Fowler A.B., Stern F., Review of Modern Physics, **54**, 437, (1982).

42. Arnold E., Surface Science, **58**, 60, (1976).

CHAPTER 3

EPITAXIAL LAYER GROWTH AND FABRICATION

3.1 Molecular Beam Epitaxy (M.B.E.)

3.1.1 Introduction

In this section, I will present an overview of the MBE process. Emphasis will be placed on the problems arising from the epitaxial growth of ternary alloy layers such as InGaAs. MBE has become the method of choice for the growth of accurately specified, crystalline heterostructures. This is particularly so for research, where low productivity is less of an issue. It was first used to create these structures in 1971[1] and has evolved into a very sophisticated deposition process, with available growth rates of between 0.1 and 10μm per hour.[2,3,4]

Unlike liquid and vapour phase epitaxy, which are quasi-equilibrium processes, crystal growth by MBE is a far-from-equilibrium process, relying on surface kinetics. The process is now so widespread that details appear in many electronic engineering and other textbooks[5,6,7,8,9] and, occasionally, whole books[10,11] are devoted to it.

3.1.2 The General Process

Elemental or molecular forms of the components that will eventually comprise the crystal structure are housed in effusion or Knudsen cells, where they are resistively heated in a high vacuum ($<10^{-10}$ Torr). An orifice can be opened allowing a collimated beam of the vapour to be directed towards the substrate, see fig. 3.1.1 below. Once the desired amount of material has been vented from the cell, the shutter can be closed. The shutter cycle can be controlled such that only one layer of atoms is accurately deposited. This layer abruptness is one of the technique's most significant features, and is ideal for the growth of low-dimensional heterostructures of the type found in this work.

The relative concentrations of the species required for incorporation into a ternary alloy layer such as $In_{0.3}Ga_{0.7}As$, can be adjusted by the control of the beam strength from each individual effusion cell. The cell-substrate pathway exists in an ultra-high vacuum ($<10^{-11}$ Torr) to prevent atomic or molecular collisions. Liquid nitrogen shrouds within the chamber help prevent cross-contamination.

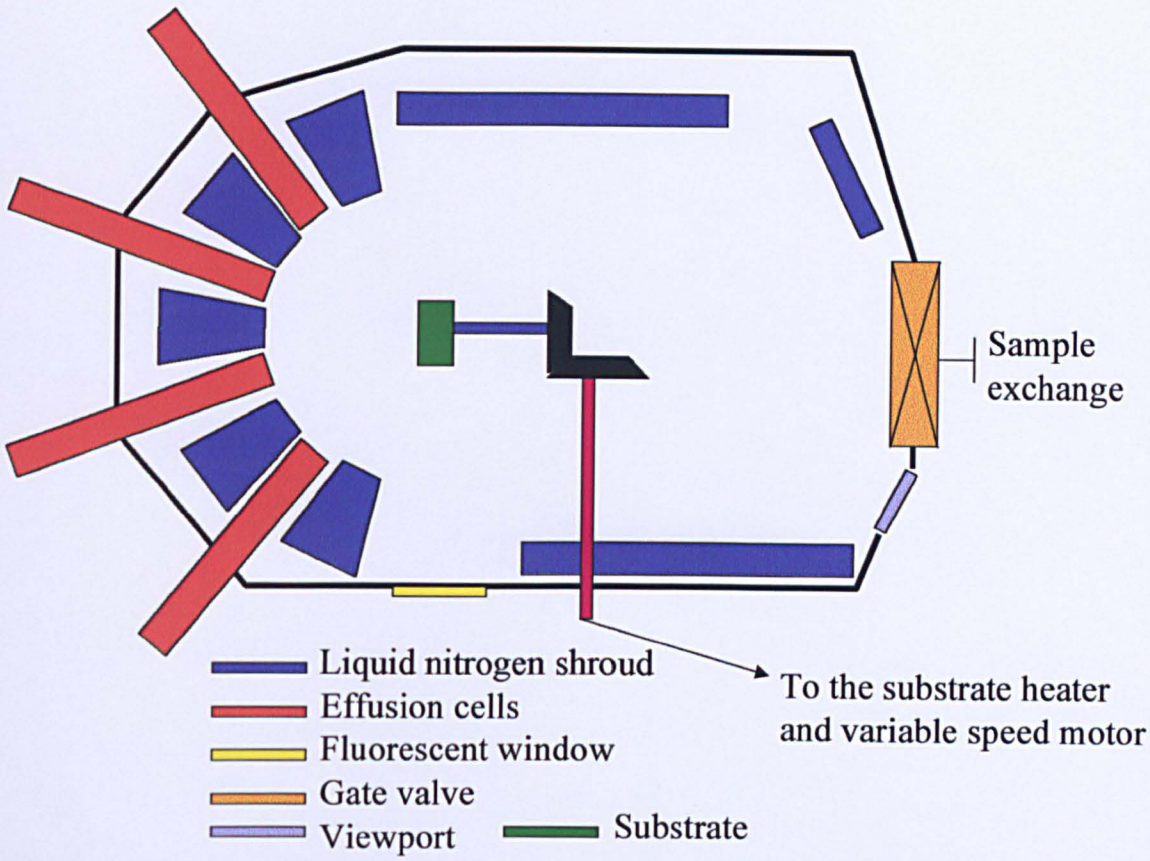


Figure 3.1.1 Schematic plan view of a typical MBE system.

The epitaxial growth of III-V compounds can be described as a series of three individual processes[12]:

a) The adsorption of the relevant atoms or molecules onto the hot substrate surface in a precursor state.

b) The kinetic surface hopping or migration of the atoms or molecules accompanied by either some absorption or dissociation or alternatively chemisorption.

c) Those species not desorbed or dissociated undergo a second order reaction and achieve nucleation and growth according to the statistics governed by the 'sticking coefficients'.

Above 500°C, arsenic atoms have a very low sticking coefficient to GaAs, unless accompanied by gallium atoms, and to counteract this, an

overpressure of arsenic is necessary. With the addition of a third or fourth species to form a ternary or quaternary alloy, sticking coefficients must be combined and the calculation of beam flux ratios can become much more complicated.

Under certain conditions of substrate temperature (>600°C), and orientation (110), the group III atoms in AlGaAs can be incorporated non-uniformly and hence cluster.[13] With a growth orientation of (100), and temperatures less than 600°C, clustering in AlGaAs is rare, but it is very common in strained InGaAs, when this compound is grown on a GaAs substrate. The tendency for an atomic species such as indium to segregate and subsequently aggregate (cluster) increases with temperature and indium concentration. The exact effect of strain on the kinetics, incorporation and ultimate distribution of the atoms involved in alloy epitaxial growth is very complicated and not fully understood.

A number of in- and ex-situ techniques are available to monitor the process[9]:

1. Reflection High Energy Electron Diffraction (RHEED) monitors substrate cleanliness and the on-going smoothness of growth. Electrons are fired at grazing incidence towards the substrate and the reflected diffraction pattern analysed. A smooth surface will produce a high-reflected intensity, whereas a rough surface will reflect poorly. The growth pattern in different directions parallel to the surface can also be monitored for growth anisotropy. A growing monolayer will therefore, produce a change of reflected intensity- starting with a high intensity, then falling to a minimum when the layer has half-formed, then rising again when the layer is complete and smooth. 'Islanding' or clustering can be detected by the

appearance of three-dimensional diffraction spots owing to the penetration of the electrons through the clusters.

2. The chemical composition of the outermost layer can be determined using Auger Electron Spectroscopy, which is particularly useful for the monitoring of dopant incorporation.
3. The above factors can also be checked using Secondary Ion Mass Spectroscopy and the ambient gas impinging on an ioniser in the chamber can be analysed using Quadrupole Mass Spectrometry.
4. Other processes can include X-ray Photoelectron Spectroscopy and Scanning Electron Microscopy.

3.1.3 The In-House MBE Growth

(Note- the author played no part in this process other than layer specification.)

For the epitaxial growth of the pseudomorphic layers, the GaAs substrates were resistively heated to a nominal temperature of 500°C or 520°C. The wafers were two-inch diameter single crystals oriented at (100). The layers were deposited in an Intervac (Varian) Gen II MBE System. In order to reduce the growth temperature, As_4 molecules were dissociated to As_2 dimers using low-temperature 'cracker' cells. These are special units that intercept the effusion beams and thermally break down the vapour molecules by collision. The As_2 dimer has a greater sticking coefficient than the As_4 molecule and may necessitate an adjustment of the beam fluxes using overpressure. Compositionally dependent growth rates,[14] which were typically about 1μm per hour,

were tightly controlled by RHEED analysis. Growth temperature was monitored by an Ircon Series V optical pyrometer.

The growth commenced with a GaAs/AlGaAs superlattice extending to about one hundred periods. This structure helps prevent the movement of defects and impurities from the GaAs substrate towards the active region above. A 600nm GaAs buffer, which reduces the likelihood that the Fermi level pins on the charge trapped in the superlattice, followed this. The remaining layers formed the active part of the device and are illustrated in fig. 3.1.2.

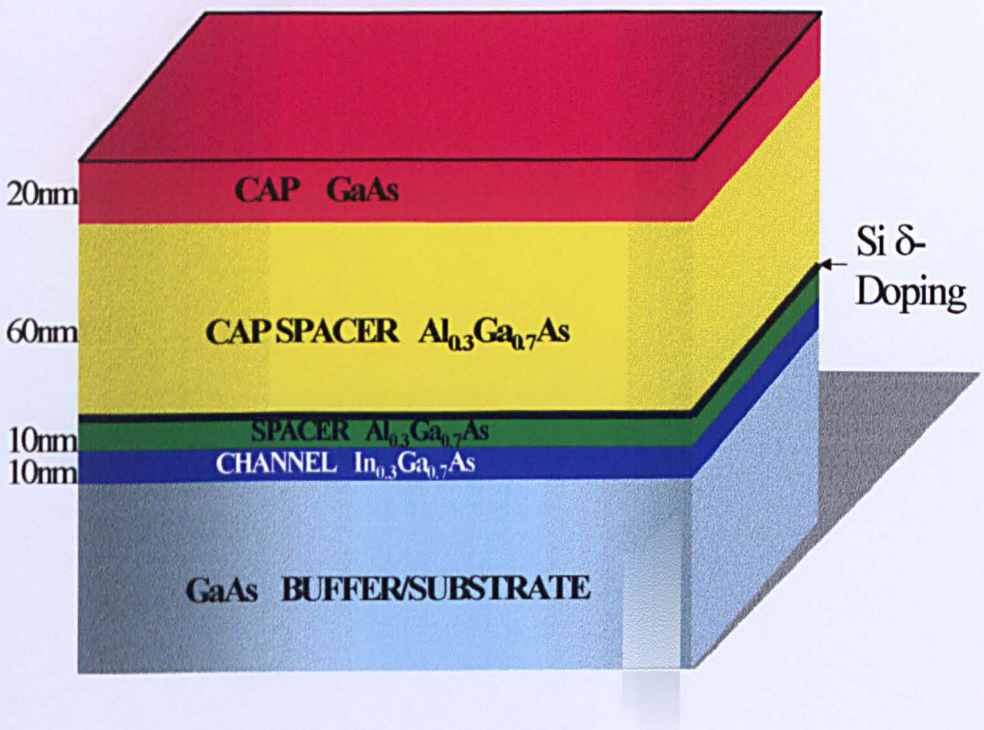


Figure 3.1.2 Cross-section of the active layers of the pseudomorphic system with a 10nm spacer and a deep channel. The GaAs monolayers are not shown. The true thickness of the substrate using the scale above would be about 7000 times that shown in the figure.

The relatively low growth temperatures used, allowed the deposition of an indium concentration of up to 30% in the 10nm wide channel layer, without apparent strain relaxation. In contrast to unstrained layers, the thermodynamics of the growth of strained layers is strongly affected, and may even be driven, by the elastic strain. People and Bean (1985),[15] calculated that a 10nm thick, 30% indium layer would not relax the elastic strain. However, in other literature,[16,17,18] this combination would relieve the strain. Such a relaxation was however, never apparent in the measured characteristics of any of the layers considered in this work. Of course, modern MBE and the low growth temperatures[19,20] used may account for this.

Next to be deposited was the $Al_{0.3}Ga_{0.7}As$ spacer layer, which offers a buffer between the channel electrons and the scattering potentials emanating from the doping layer. The width of this layer also influences the electron concentration in the channel. Also affecting this is the donor charge, which was provided by the planar or delta-doped silicon (Si), and was sandwiched between five monolayers of GaAs. The lower three monolayers act as an impurity trap during the subsequent Si deposition. These three layers, along with the upper two monolayers, prevent thermal over-diffusion of the Si. To allow for the switch to Si deposition, an interruption in the growth must occur. This time delay may also produce aggregation at this interface of some indium metal, which has segregated and floated upwards during the previous growth.[21-28] An indium-rough interface may ensue and this possibility should be considered in any analysis.

The remaining layers, which are undoped, were chiefly incorporated to provide a semi-insulating spacer between the Schottky gate and the doping layer. The cap offers additional favourable

characteristics for the deposition and operation of a Schottky gate metal. This upper layer configuration was designed with simple characterisation in mind, but the layers used by Glasgow University's Nanoresearch Centre, for their ultra-fast integrated circuits, differ around the cap layer. The gates for these devices are recessed into the cap layer, therefore, an etch-stop layer of 5nm of AlGaAs is added. An additional surface buffer layer (2.5nm GaAs) prevents device degradation due to the subsequent oxidation of the AlGaAs upper spacer, which is kept thin to reduce field fringing from short (about 100nm) gates. As a major aim of this work was to characterise material for these devices, it was important to keep this group's requirements in mind.

3.2 Cleanroom Fabrication

3.2.1 Introduction

The original intention was to employ photolithographic masks for defining the Hallbars and indeed, they were used for the fabrication of the first sample (A610). Although the smallest critical dimension existing on the Hallbar is about 20μm, it was perceived that electron-beam (e-beam) lithography would produce sharper corners and cleaner edges than those apparent after photolithography. Any such rough features might conceivably affect the measured device characteristics,

and render the analysis more confusing. Therefore, all subsequent wafers were fabricated using e-beam lithography.

Wafer cleaning and other processes were carried out within a cleanroom environment with at least a rating of 100. This implies a maximum of 100 particles (>0.5μm) per ft^3. The e-beam lithography stage was conducted within a room having a higher rating of 10 (≤10 particles per ft^3), owing to its general use for patterning sub-micron devices. To prevent further contamination, a cleanroom suit, shoes and gloves were always worn during device fabrication. The latter prevent dirt, oils and salts being spread from the hands, but also offer the skin protection from chemicals.

3.2.2 Wafer Preparation- Cleaning

Cleaning involves the removal of undesired environmental or chemical material that persists on the surface of a wafer, subsequent to a given step in the fabrication process. It is crucial for attaining high yields and reproducibility, and can badly affect any of the implemented stages.[29] Prior to cleaning, any loose particles attached to the wafer surface were blown off using a flow of nitrogen (N_2) gas. Residual resist, grease, embedded particles, etc., accumulated from the previous fabrication stage or the MBE stage, were removed using highly pure organic solvents. Of these, residual resist was the most troublesome organic contaminant. This removal of resist and grease was accomplished in two stages.

Each quarter wafer segment, cleaved from a two-inch diameter whole, was subjected to successive soaks of acetone and iso-propyl

alcohol (IPA or propan-2-ol) in clean plastic beakers. The soakings were implemented at room temperature - each lasting five minutes, and were greatly assisted by the application of ultrasound in an ultrasonic bath. The wafers were rinsed using IPA, as it is relatively stable in air and does not leave prominent drying marks. Drying was accomplished swiftly using a flow of nitrogen gas. Other more powerful, but potentially damaging methods were available, such as wet acid immersion, or oxygen plasma-assisted oxide removal (ashing), but were never required. Organic solvents are ideal in that their reaction selectivity leaves the wafer and the intended depositions unaltered. However, acetone with the addition of ultrasound, can destroy unannealed metal deposits, for example, gate metal. Therefore, care was taken as to when and how acetone was applied. An overview of the technology of wafer cleaning is contained in a chapter by Chang and Chao.[30]

3.2.3 The Chemical and Physical Properties of E-Beam Resist

3.2.3.1 Resist Spinning

(In this section, the description of resists will be restricted to that applicable to the electron beam lithography process.)

After cleaning, the next process involved the application of one or more layers of liquid resist to the quarter-wafer by 'spinning'. The wafer

was held on a vacuum chuck and rotated, along with the film of resist, until a terminal angular velocity was achieved. To obtain the desired film thickness, this velocity must be applied for an appropriate time. The final thickness must neither be too thin nor too thick. Resist that is too thin could result in damage to the surface of the wafer when the subsequent beam writer dose is applied. Additional problems arising from lateral exposure of the resist may also ensue from electrons back scattering from the semiconductor. This is the so-called Proximity effect, and is one of the major practical limitations to the minimum size of pattern features. A resist layer that is too thick will not be fully penetrated by the electron beam and hence the exposed pattern will not be properly defined on the wafer surface. Therefore, the spin time and speed were set according to the number and type of resist layers applied and the subsequent e-beam dose and energy.

For all of this work, two layers of different composition were applied separately by plastic pipette to form a bi-layer. The reasons for this, including the chemicals chosen, will be described later. The first, or lower layer, was composed of 15% Aldridge® resist or BDH® resist in chlorobenzene, spun at 5000 r.p.m. for 60 seconds. Any liquid resist clinging to the underside of the wafer was removed with acetone to prevent alignment difficulties during electron beam writing. The wafer plus resist film was then placed on a watchglass, covered with a petri-dish and, when Aldridge was used, baked at 180°C for 45 minutes. When BDH was used the settings applicable were 180°C for 120 minutes. The Aldridge tended to produce a residual scum after lift-off if the baking time exceeded about 45 minutes. On completion of the baking and cooling, the second layer was spun at the same speed and for the same time as before. It was always composed of 4% Elvacite® in

xylene, and was baked at 180°C for about 140 minutes. After cooling, the resist was now ready for electron exposure.

3.2.3.2 Electron Beamwriting Lithography

The resist was exposed to electrons emitted by a Leica EBPG5. All the parameters required for its operation could be set using software.

Software

Before patterning of the resist by e-beam exposure can commence, the parameters defining its operation must be set and transferred to the electron beam machine. There were a number of software processes involved, from the initial computer-aided-design (CAD) to the final computer-controlled e-beam operation. The flow of data between these two end-points is summarised in fig. 3.2.1. The design of the pattern that would eventually be defined by the electron exposure was set in the WAM® C.A.D. software package, within an I.B.M.-compatible P.C. For the Hallbar/ field effect transistor devices in this work, see fig. 3.2.2(a-d), three individual layers, the ohmic layer (red), the isolation layer (green) and the gate layer (gold), had to be individually designed (fig. 3.2.2d explains why the gate metal does not, despite appearances, short-circuit with the channel). They had also to be accessible as individual patterns from within the overall pattern. Therefore, in WAM, each of these sub-patterns was defined as a separate layer. This was necessary as each layer had to be passed

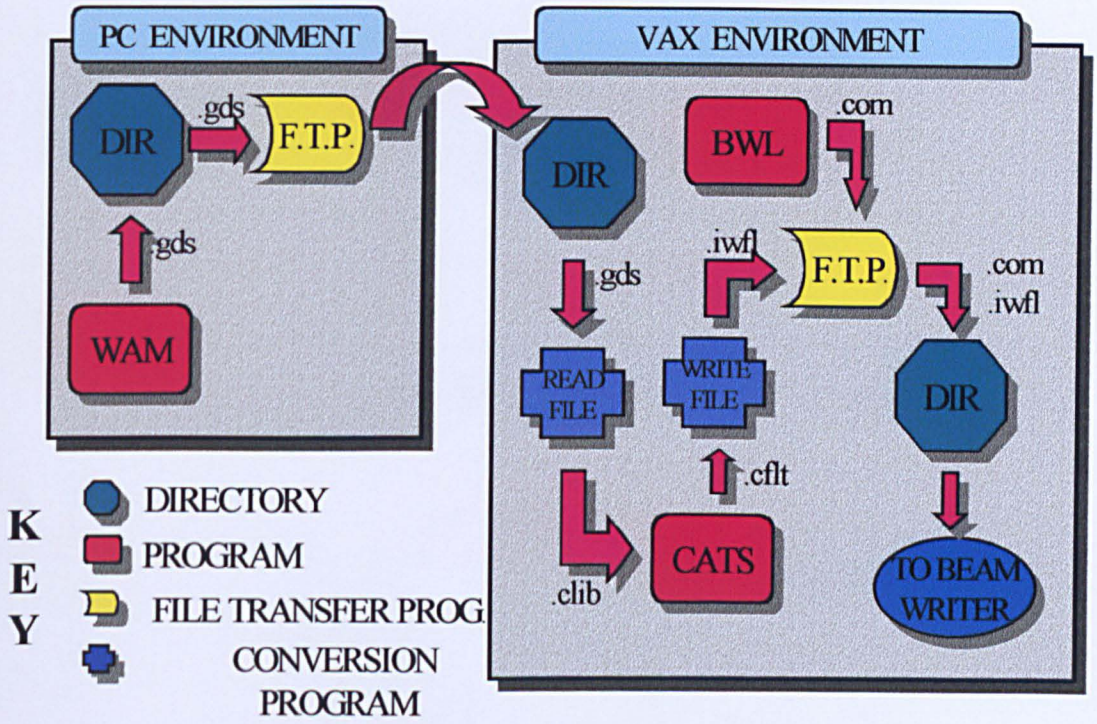


Figure 3.2.1 Schematic displaying the flow of data from the design software (WAM), through that involved in parameter setting for the beam writer operation (CATS, BWL).

separately through the whole cleanroom and lithographic process.

The standard CAD format file (.gds), containing the WAM CAD data for the basic device (fig. 3.2.2), was then sent by file transfer protocol (FTP) to the UNIX-based VAX system, which contains the remaining software. The READFILE conversion program translated the .gds file into a standard library format file (.clib) readable by the next program- CATS. In response to user inputs such as the basic device pattern file and name; the sub-layer within that pattern to be 'fragmented' or isolated; the required resolution of the e-beam spot, etc.,

CATS produced an output data file (.cflt). This file was translated to a file (.iwfl) suitable for processing by the final beam writer software using a CATS sub-program - CATS WRITEFILE.

To complement this file, another program in the VAX system, BWL, provided the rest of the required data in response to user inputs. While CATS deals mainly with the shape of the pattern, BWL is more

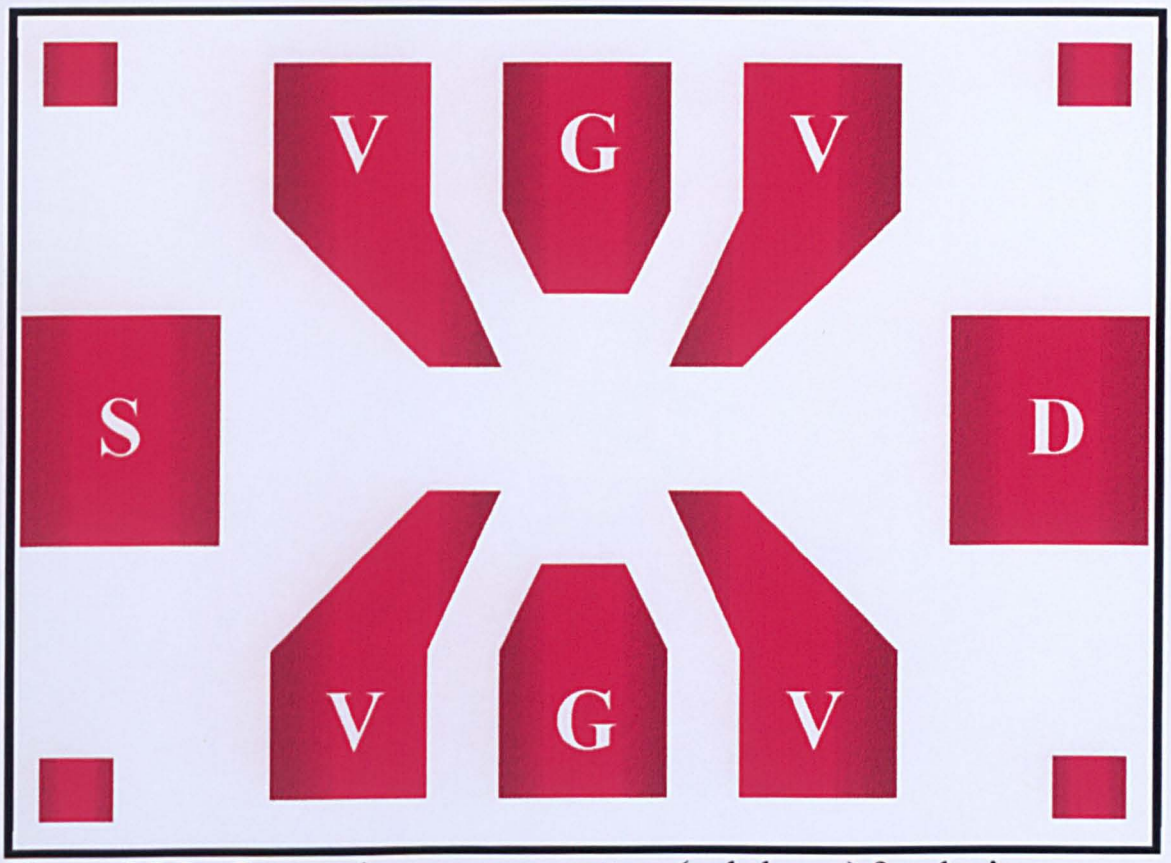


Figure 3.2.2a Beamwriter exposure pattern (red shapes) for ohmic layer design. *S* is the source ohmic, *D* is the drain ohmic, *V* are voltage measuring ohmics, *G* are gate ohmics and the other 4 small squares were used for testing the subsequent isolation, see figs. 3.2.2b and c. (The markers for registering the future patterns are not shown.)

concerned with its distribution over the supplied wafer, and the parameters required for correct electron exposure of the overlying resist. Both the isolation and gate lithographic cycles require extra information regarding their pattern placement with respect to the ohmic layer

pattern. This can also be set in BWL by defining their position relative to a special 'markers' pattern, which was incorporated as part of the original ohmic layer. Strictly this is not recommended as the subsequent annealing of the ohmic metal can render these markers less visible to the beam writer detectors. However, this never proved a problem and did save the considerable time and expense of a full 'markers' cleanroom and lithographic cycle.

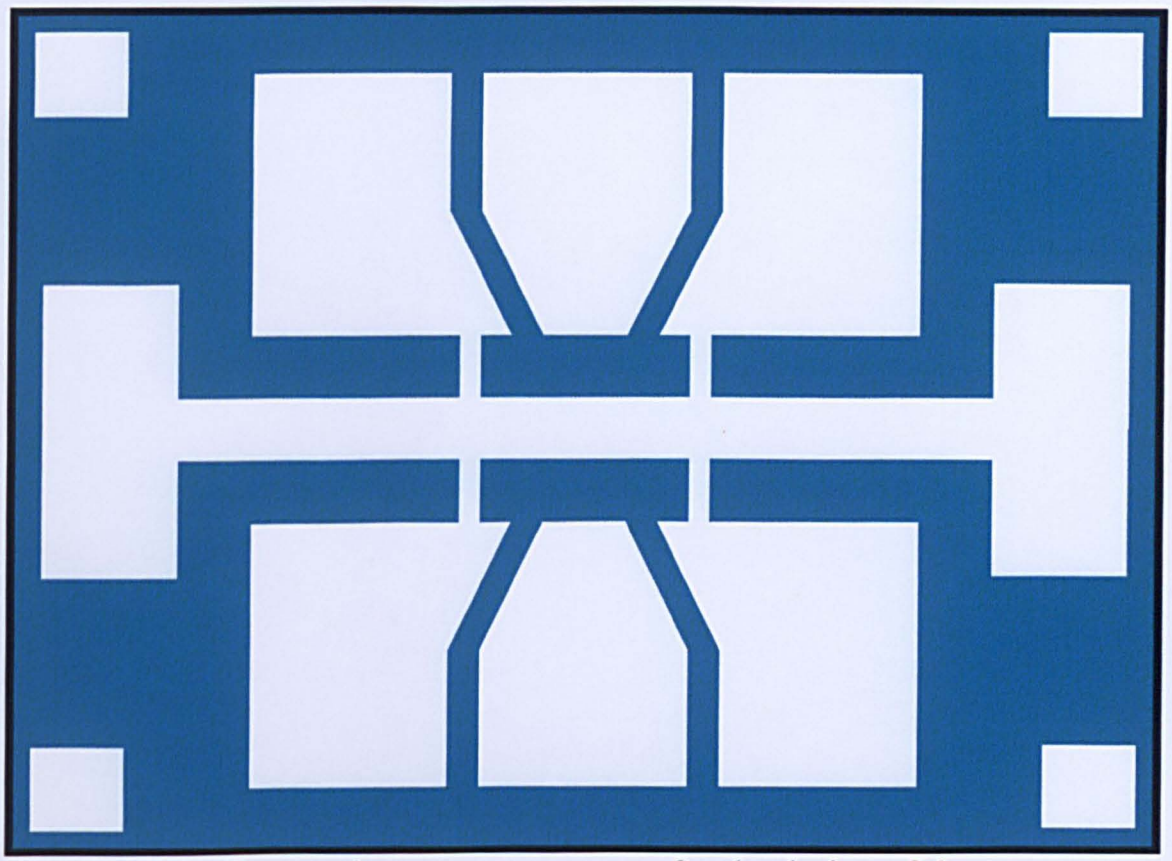


Figure 3.2.2b Beamwriter exposure pattern for the design of the isolation layer. The green area was the area exposed to the electrons. Bear in mind that the annealed ohmic patterned metal exists on the wafer at this point (see fig. 3.2.2c).

The output BWL file (.com) along with the CATS WRITEFILE output file (.iwfl) were both transferred in this form to the beam writer directory, where the operators could access and load it for beam writer

operation. The important parameters input with regard to the pattern size and the selected resist type were:

1 Electron beam voltage- 50kV

2 Resolution- 0.1μm- (The accuracy to which the beam can be placed).

3 Dose- 300μCcm^{-2}

4 Spotsize- 160nm

E-Beam Operation

The electron beam machine operates by directing a beam of electrons with a Gaussian intensity profile towards the resist using a sophisticated variable magnetic field. The use of a Gaussian profile permits a very small spot size (minimum 12nm, maximum 400nm) for high-resolution work. The beam scans serially across the appropriate areas of the resist as defined by the pattern software data. The other three parameters applied to the beam are the voltage (minimum 50kV, maximum 100kV); the dosage (minimum 1μCcm^{-2}, maximum 300μCcm^{-2}); and the resolution (minimum 5nm, maximum 312.5nm). The first defines the energy (E) and hence, the momentum (p) acquired by each individual electron. Using de Broglie's formula, $\lambda=h/p$ (where λ is the wavelength and h is Planck's constant) the implication is that the electron wavelength will decrease with increasing momentum. High resolution will then be theoretically available. For example, if 50kV is applied to the electrons, and using an alternative version of the formula

above, $\lambda=h/(2Em)^{1/2}$ (where m is the electron's mass), then the wavelength obtained is 5.5pm. The practical resolution will not be

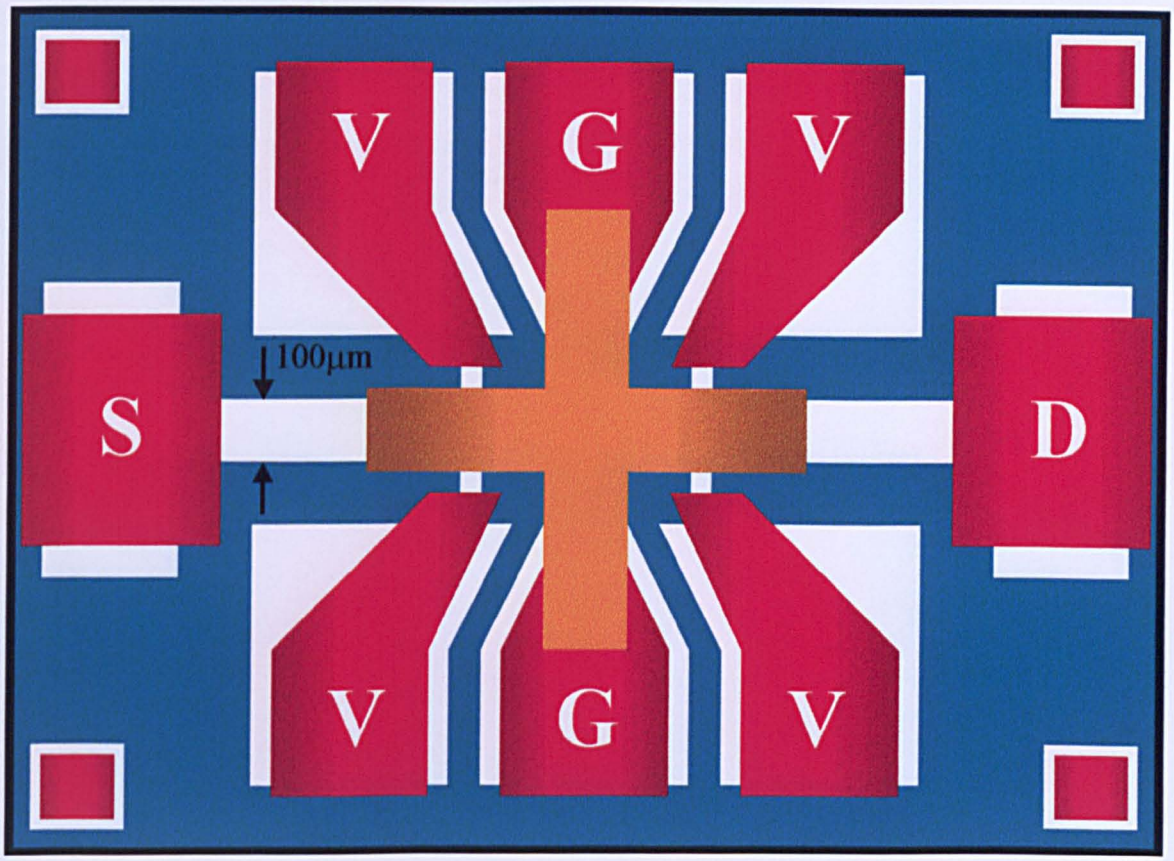


Figure 3.2.2c The complete Hall bar design. The registration between the ohmic layer (red), and the isolation layer (green) can now be seen. The gate layer (yellow) has also been added. Note how the gate and the small square isolation ohmics are, ideally, entirely electrically isolated from the Hall bar, but that the current and voltage ohmics have direct access to it. The letters are explained in fig. 3.2.2a.

nearly as good as that for a number of reasons, including aberrations. The second parameter, the dosage, determines the number of electrons and hence the charge, which strikes the surface of the resist per unit area per second. The third parameter, the resolution, describes the accuracy

with which the beam can be placed, and is usually similar to the scan step size.

Figure 3.2.2d Schematic cross-section of the Hallbar showing how the etch isolates the channel from other regions. At first sight, the gate metal appears to short-circuit with the channel, however this does not happen. Electrons occupying the surface states in this region are non-conducting as the states are mid-bandgap in energy. The result is an insulating layer.

To select values for these parameters, it was found prudent to fix the beam energy (50kV); the spot size (160nm); and the resolution (100nm) appropriate to the smallest feature size (20μm) found within the given layer design, and only vary the dosage in a series of exposure tests. A useful 'rule of thumb' for setting the fixed values above is that the spot size should be about twice the resolution, which itself, should be at most $1/10^{th}$ of the minimum feature size. The best value found for the dosage was $300\mu Ccm^{-2}$. The smallest feature size did not warrant the application of the highest beam voltage/energy available on the beam writer; higher resolutions or smaller spot sizes, and therefore, the exposure time saving was significant. The isolation layer, which required the greatest area to be exposed, only needed about 3-4 hours for a matrix of about 70 patterns. This exposure time is kept down by the use of vector scanning, whereby only the areas to be exposed are scanned. Raster scanning, on the other hand, would involve scanning the whole surface area whilst the beam is 'blanked' and 'unblanked' over

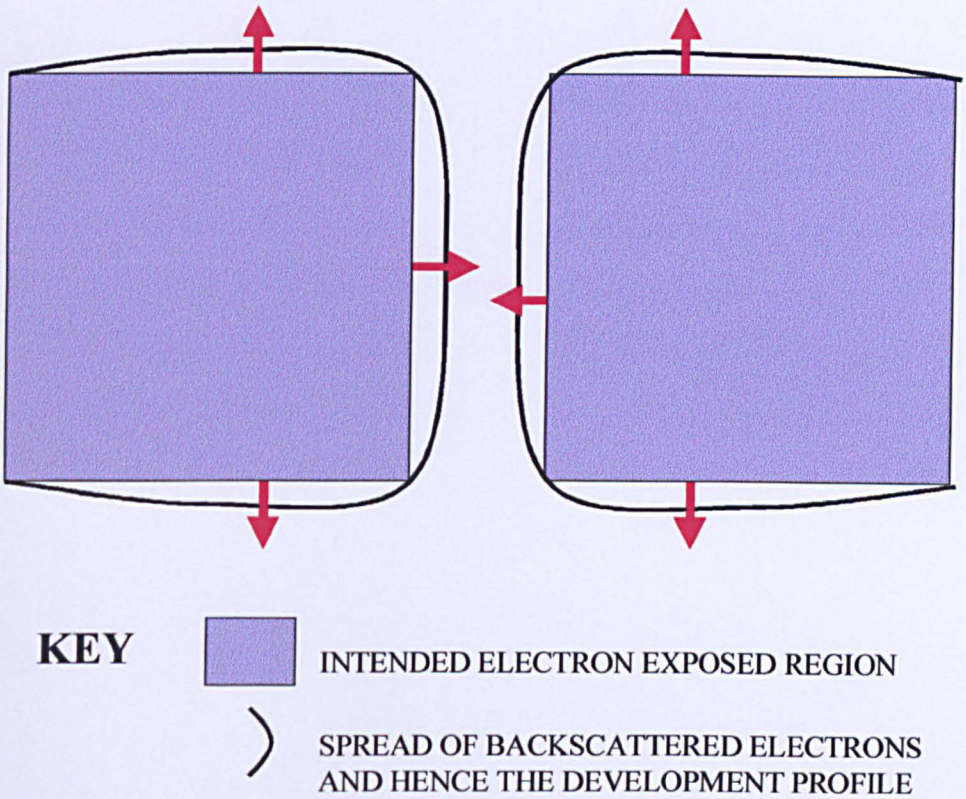


Figure 3.2.3 Schematic diagram illustrating the Proximity Effect. Note the lateral spread of backscattered or secondary electrons and its effect in rounding corners of the developed profile, and narrowing unexposed channels (about 100nm in figure). Over etching can have its own type of lateral effect and may even pinch-off channels.

the appropriate areas- analogous to the operation of an inkjet printer. Of course, one disadvantage of both vector and raster scanning is its serial application and thus, its implications on productive yield. It is much slower than the alternative parallel method of photolithography, where ultra-violet (UV) radiation is applied through a mask for only about 12 seconds. However, this method is limited by the relatively poor resolution engendered by the UV wavelength. Whilst photolithography was suitable for producing the smallest feature size in this work, the

edges and corners of the Hall bar appeared much rougher and less well-defined than those produced by e-beam lithography.

As stated earlier, it was important to apply the correct dosage for a given resist strategy, as an overdose would precipitate a strong Proximity Effect. This effect can even be induced by the 'correct' dosage, as the electrons are exposed to a range of energy-reducing forces within the resist. The higher energy electrons can penetrate about 5μm into the underlying semiconductor, whereupon lateral back-scattering, to a radius of up to about 2-5μm, can occur. This effect can 'round' corners and narrow unexposed channels, as shown in fig.3.2.3.

To avoid an excessive angular deflection of the electron beam, and hence, distortion of the spot, the beam is scanned over small block areas or 'frames'. To address a neighbouring frame, the stage supporting the wafer is translated appropriately. The frames are then 'stitched' together with some unavoidable overlap (40nm). Therefore, if sub-micron features existed, it would be desirable to isolate them from these stitching edges.

Electron-Beam Resist Chemistry

The electrons chemically alter the region exposed in two very different ways. Resist materials utilized in the e-beam process are very often based on organic polymers, for example, polymethyl methacrylate (PMMA), particularly if a lift-off is required, see later. "Resist" acquires its name from its ability to shield the underlying material from subsequent processing steps, such as etching, metal deposition, etc. Resists exist in two forms- 'positive' or 'negative', and are fairly similar

in that they contain the same basic 'building blocks' of PMMA monomers:[31]

$$-[CH_2\text{-}CCH_3(COO(CH_3))]-$$

However, they differ in the active components added to the monomers, such as styrene, epoxy and vinyl groups.

On entering the resist, a high-energy electron undergoes electro-magnetic interactions with the electrons of the resist atoms, resulting in a rapid loss of energy. This energy produces either the promotion or the inhibition of monomer cross-linking, depending on the resist type. In a negative resist, chemical cross-linking between the molecules is encouraged on exposure to an electron beam. This renders the polymer far less soluble to future development, with only one cross-link per molecular pair being necessary. In contrast, in a positive resist, cross-links are discouraged, which implies that those regions exposed to the electrons will be subsequently far more soluble in the appropriate developer. It should be emphasised that although one 'direction' of the above chemical reactions is favoured, the other 'direction' will exist to a much lesser extent. Therefore, to avoid the unwanted reaction, the dissolving developer should be rinsed off quickly after use. After development of a negative resist layer, the regions exposed to the electrons will remain, and hence, shield the underlying semiconductor. In contrast, those same exposed regions of a positive resist will, on development, dissolve away and no longer shield the underlying areas. The unexposed resist now undertakes that role.

For a number of reasons,[9,31] positive resist is generally more popular, particularly for e-beam exposure. These include:

1. Better contrast, edge profiles and line-width control.

2. There is less swelling on exposure to development solvents.
3. It is much better for processes involving lift-off.
4. It is easier to remove, post-processing.
5. In general, it consumes less exposure time, as 'device' areas on a wafer are usually smaller than the remaining area.

The resists used in this work were always positive.

3.2.3.3 Development

Developers are solvents which can dissolve those regions of a resist where cross-linking has been inhibited. They are organically based and, for a given type, vary in the relative proportions of active ingredients. This is to permit effective development of resists, which themselves contain varying active proportions.

In order to dissolve the resist, the beam-printed quarter-wafer was submerged in a 50ml plastic beaker full of the mixture- (1:1) methyl-iso-butyl-ketone combined with iso-propyl-alcohol, (MIBK/IPA). Of the available mixtures of these two chemicals, this produced the fastest development times, but the lowest resolution and contrast. The temperature of the mixture was 24°C, and the submersion lasted 60 seconds. IPA was used as the rinse, which, as stated earlier, must be effected rapidly to avoid over-development. After drying with a stream of N_2 gas, the developed patterns were examined in a light microscope.

After this stage, the relevant fabrication processes, metallisation, etching, etc., would be implemented, but descriptions of these will be presented later. One process, 'lift-off', which relies heavily on resist characteristics, still remains to be detailed.

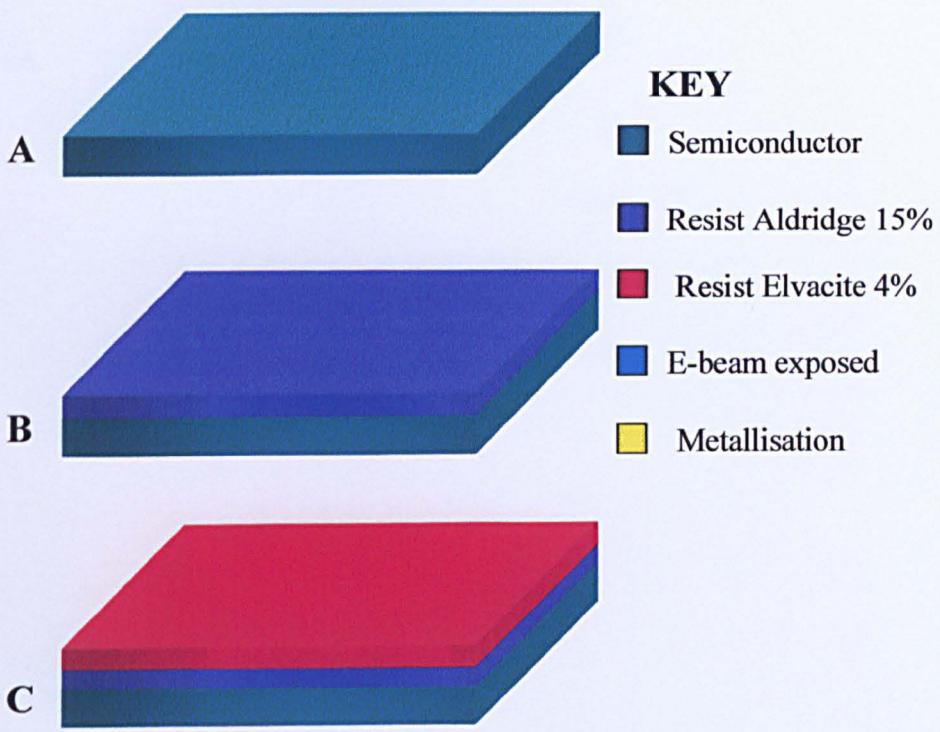
3.2.3.4 Lift-off

Much of the technology invested in resists/ developers was used to facilitate the so-called 'lift-off' procedure. It is a process whereby unwanted material depositions can be carefully removed by dissolving slowly and carefully the underlying resist. The material deposition is usually a metallisation and the resist, of course, is that left undissolved by the previous development. Therefore, the solvent must be very different from the developer in its capability to dissolve unexposed positive resist. The difficult function, however, is the clean detachment of the metal overlying the boundary edges of the resist. One of the best methods to achieve this is described below, and was used in this work.

The sequence of events describing the production of a metal layer, for example, the ohmic layer is shown in fig. 3.2.4. The two layers of resist are separately spun onto the semiconductor, see fig. 3.2.4(A-C). The lighter Aldridge (molecular weight- 120,000 a.m.u.) but higher concentration (15%), resist forms the lower layer. The heavier Elvacite, (molecular weight- 350,000 a.m.u.) but lower concentration (4%), forms the upper layer. In the figure, both layers are shown as the same thickness, but in reality the lower was about 14 times that of the upper - about 1400nm : 100nm. Fig. 3.2.4(D-F) shows the sequence of exposure, development and metallisation, respectively. Again, the thickness of the resist is much exaggerated relative to the yellow metal layer (about 600nm for ohmics), and the apparent 'cliff-face' does not exist. This bi-layer resist sequence permits, on development, an undercut to establish in the lower layer, see fig. 3.2.4G. As mentioned earlier, this aids a clean cut in the metal film at the resist junctions, and thus prevents the formation of ragged metal edges in the final product - fig.

3.2.4H. The lift-off itself, was carried out in a beaker of warm acetone at 45°C, which dissolved the unexposed resist over a period of about one hour, thus easing the upward removal of the overlying metal.

Many multi-layer techniques have been researched to achieve a good lift-off or to reduce the effect of the back-scattered electrons during the exposure stage. Sometimes, each layer requires separate development. For examples, see Lin and Gozdz[32] (1988), Williams[29] (1990), Hatzakis[33,34] (1981, 1984) and Oro and Wolfe[35] (1982).



/continued over

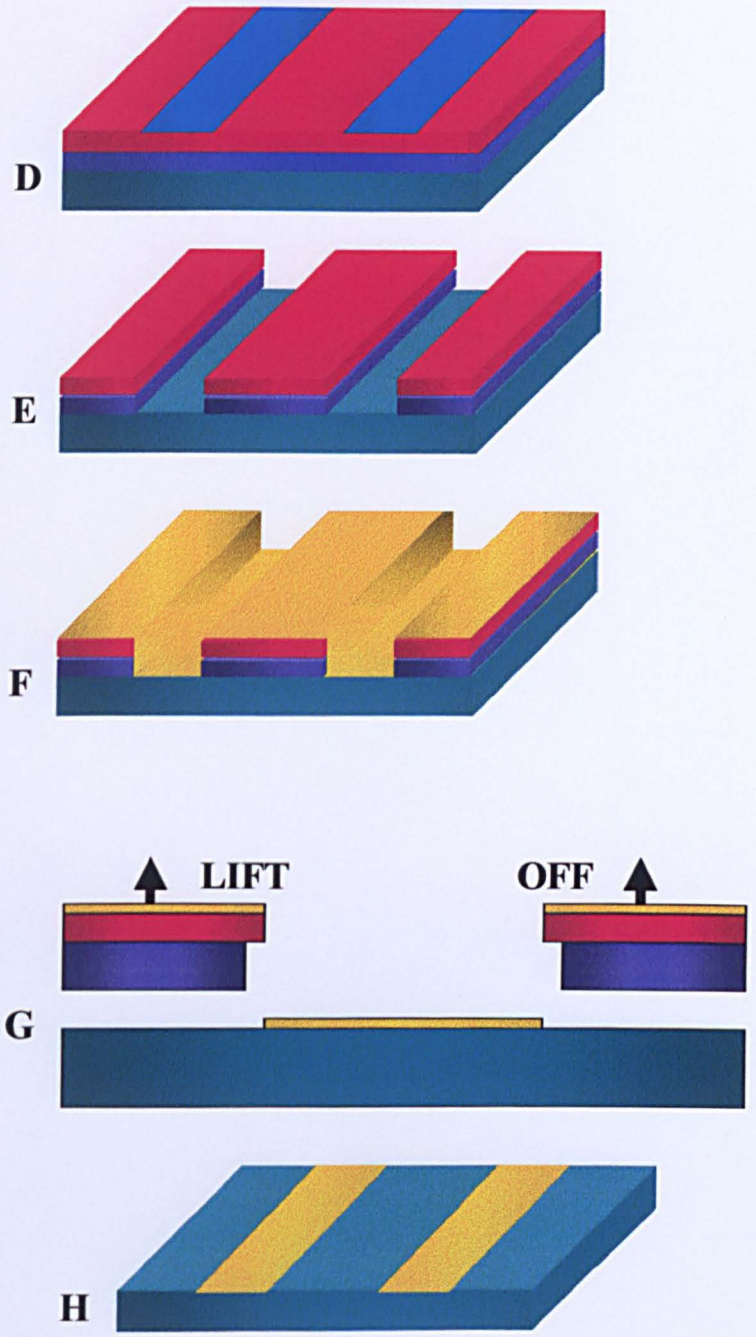


Figure 3.2.4(A-H) Schematic representing the sequence of stages involved in depositing metal shapes on a semiconductor surface. **A** is the cleaned wafer. **B** has the addition of the lower resist layer- 15% Aldridge. **C** has the addition of the upper resist layer- 4% Elvacite. (relatively much thinner than shown) **D** shows the surface areas exposed by the beam-writer. **E** displays the corresponding bulk regions removed after development. **F** shows the surface after a metallisation evaporation. **G** demonstrates the lift-off removing the unwanted metal. **H** exhibits the final desired metal pattern.

3.3 Metallisation

Metallisation is the deposition by evaporation, sputtering, etc., of a metal or mixture of metals onto a surface. For device fabrication, the most common process is evaporation of a metal mixture onto a semiconductor surface to form 'gates' or 'ohmic' contacts. Both these end-points demand specifically defined shapes on the semiconductor surface, and hence also appropriately exposed and developed resist patterns, as described in section 3.2.3.3. As stated there, shaped metallisation requires a lift-off process to remove the unwanted metal. The two processes which produce ohmic and Schottky gate contacts will be described separately, and in their fabrication sequence.

3.3.1 Ohmic Contact Metallisation

Before the presentation of the quarter-wafer for evaporation, it was cleaned, spun with resist, written by the electron beam and developed, as described in sections 3.2.2 and 3.2.3. However, one more process was required. In order to remove oxides, which will grow on the wafer surface now exposed after development, the wafer was submerged in a de-oxidising solution. This consisted of a mixture of 4 parts (80ml) of reverse osmosis (RO) water to 1 part (20ml) of concentrated hydrochloric acid (HCl), with 1ml of 'FC93' solution added to aid wetting of the surface. The submersion lasted 30 seconds at room temperature, within a 100ml plastic beaker dedicated to this process to reduce contamination. An ohmic contact should, of course, have as little

resistance as possible across its junction therefore, the presence of an oxide film could degrade its performance.

The quarter-wafer was rapidly loaded into a special holder, which when loaded into the evaporator exposed the surfaces downward to the metallic beam. The evaporator used was a Plassys MEB 450, which offered two chambers for housing the holder. The first- the transfer chamber- permitted both the loading of the holder, and when under a vacuum of less than 10^{-4} bar, the transfer of the holder to the evaporating chamber. When it was pressurised to less than $2x10^{-6}$ bar, deposition of the appropriate metal mixture could begin.

Each metal was evaporated separately in the following order during a process requiring about one hour. The software recipe which initiates this sequence is 'Ohmic 7'.

1 Pump down and clean.
2 Nickel- 8nm
3 Germanium- 120nm
4 Gold- 130nm
5 Nickel- 80nm
6 Gold- 250nm

After venting and unloading, the wafer was transferred to a beaker of warm (45°C) acetone for lift-off. Lift-off was always successful, as the patterns were quite large.

3.3.2 Annealing

To produce ohmic contacts, alloying and annealing were required, in order to allow electrical contact with the 2DEG, sited below in the channel layer. To achieve this, the metal underwent rapid thermal annealing (RTA) with the assistance of plasma to keep temperatures as low as possible, see fig. 3.3.1. In this process, some of the metal mixture is driven down from the surface in a complicated process, whereby some metals diffuse before others and to different depths. Care must be taken that the chosen annealing cycle was adequate for channel access by diffusion.

Annealing 'recipe 1' was chosen, which contains a rapid temperature rise to 360°C, with a slight pause at 310°C. The terminal temperature was applied for 60 seconds, and was controlled using a thermocouple, which made contact to the base of a silicon wafer

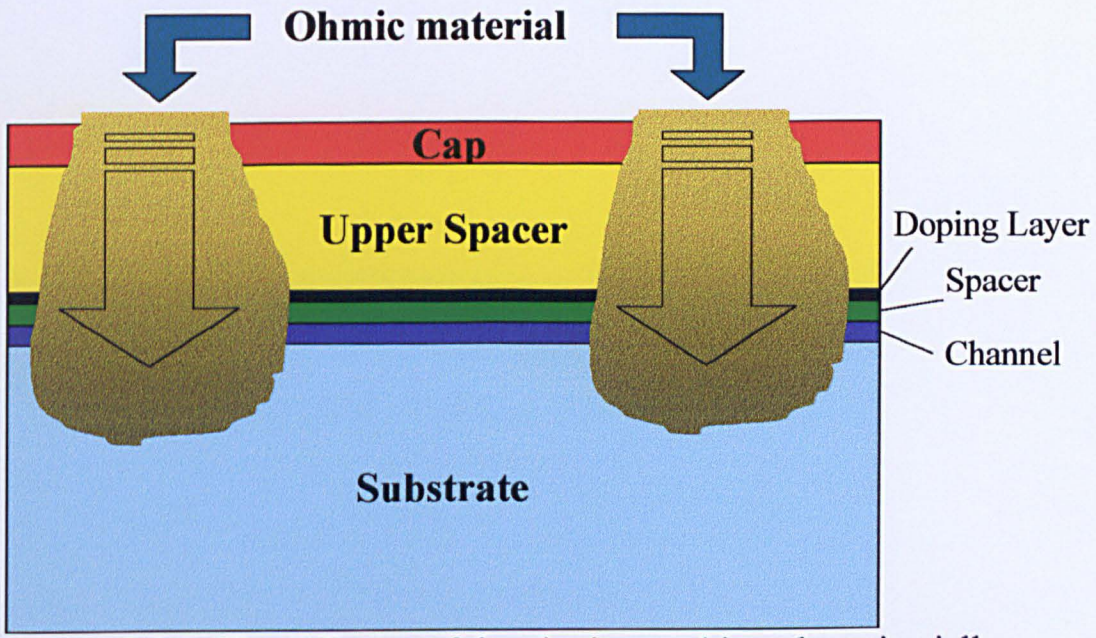


Figure 3.3.1 The diffusion of the ohmic metal into the epitaxially grown layers of the wafer during the annealing process. Note that the channel layer has been accessed.

susceptor or sample platform. The temperature variation was monitored by a chart recorder system. It was extremely important that a clean, and distortion-free, susceptor was used, otherwise temperature control was much less efficient, and resulted in the ohmics having a 'bubbly' appearance and texture. The success, or otherwise, of the annealing was checked by two-terminal DC resistance measurements at room temperature. Typical values for resistance in structures of these dimensions and channel type were about several hundred ohms after successful anneals.

3.3.3 Gate Contact Metallisation

The procedure for the deposition of a Schottky gate metal was similar to that described in the previous section, except that no de-oxidation of the exposed wafer surface was necessary. A high resistance contact is desirable here, therefore the removal of oxides was unnecessary. The other major differences was in the recipe used -'Gates 1' and, of course, the absence of annealing. 'Gates 1' produced the following deposition sequence, which lasted about 20 minutes:

1 Pump down, no clean.
2 Titanium- 33nm.
3 Gold- 160nm.

The lift-off was identical to that implemented for the ohmics, but as the metal was not annealed, care had to be taken with any subsequent cleaning, as the gate metal was very fragile. After any lift-off procedure, it was difficult to be certain of the complete removal of all the resist,

despite the long submersion in warm acetone. After the ohmic or etching stage, the wafer does of course, undergo cleaning, including the application of ultrasound. This does remove all the residual resist, but cannot be applied after a gate deposition, as the metal will be broken-up and dispersed. Therefore, some very thin residual resist may exist on the surface of the ohmic pads, and its possible presence should be borne in mind.

3.4 Etching

Etchants generally exist in two forms- 'wet' or 'dry'. A wet etchant is a mixture of several solutions in RO water, each with a tightly controlled concentration. If these concentrations are accurately applied, then the rate of etching into a given material can usually be relied upon. A typical wet etchant rate into semiconductor material is 1 or 2 nm per second, but variation is possible. Whilst the application of wet etching is relatively quick and simple, two major disadvantages exist with its use:

1 Significant lateral dissolution of the semiconductor can occur, that is, sideways etching of the material that lies underneath the resist. The effect is analogous to the proximity effect, when electrons are back scattered, see fig. 3.2.5. It is more likely to happen when the active ingredient concentration(s) and/or etching time are too large. A possible result would include the narrowing of channels as defined by the etch-pattern, for example in a Hallbar. The much narrower channels branching off towards the voltage measuring ohmics, see fig. 3.2.2c, were particularly susceptible to this effect, and this indeed, occurred once. As can be seen in fig.

3.2.2b, if lateral dissolution took place, the etched areas (green colour) would gradually meet either side of one of these narrow channels, thus blocking the intervening 2DEG pathway.

2 The other problem can be the roughness of the edges remaining after contact with the etchant. This phenomenon can promote additional electron scattering factors, thus complicating the analysis of the wafer's intrinsic scattering mechanisms.

To diminish these problems somewhat, dry etching can be introduced. This involves active molecules in the form of a plasma being allowed to strike the exposed surfaces (resist plays a similar role here as in defining the regions to be wet-etched), and hence, etching down by mechanical abrasion and chemical reaction. This produces a much smoother, vertical etched profile, which may be essential for the satisfactory analysis of quantum-sized structures, such as quantum wires and lateral superlattices. Etch rates vary with the species and conditions applied, but are roughly comparable with the wet etch rates. Dry etching was never required in this work as the fabricated structures were relatively large and diffusive in transport character. Etching was only necessary for the definition of the measuring Hallbars, with the major concerns being the isolation of the 2DEG in the appropriate regions, and preventing any significant lateral dissolution.

For the isolating etch, the wafer plus developed resist, was submerged in a 100ml plastic beaker containing a thoroughly mixed solution of 100ml of RO water; 1ml of 30% hydrogen peroxide (H_2O_2); 1ml of ammonia solution (NH_3); along with 0.7ml of the wetting agent FC93, which ensured the etch attacked the surface uniformly. This solution etched at a rate of 1nm per second, and the total submersion time therefore, depended on the channel depth in the wafer. After

rinsing with IPA and drying with N_2, the etched trenches were checked using a light microscope. Only once was over-etching apparent with no under-etch, that is, non-isolation of the Hallbar, mainly because the etching was allowed to continue a further 10nm below the channel and into the substrate.

Although strictly unnecessary, as no lift-off was required, a resist bi-layer was spun for this fabrication stage. This permitted the same beam-writer dosage to be applied for all stages to foster step uniformity, although the absence of the much thinner top-most resist layer would not make a great deal of difference.

As a final measure, the isolation was electrically checked using two-terminal, DC current/ resistance measurements at room temperature. Probes were placed on the sample usually between the special, isolated small ohmic pads found at the four corners of the design, see figs. 3.2.2a and c. Typical maximum currents obtained were a few nA in the light, and several pA in the dark. These values were entirely acceptable. Currents were noted, as resistance was impossible to infer from the noisy appearance of the I-V data, which of course, was to be expected across isolated regions. The resistance along the Hallbars, between various ohmic pads connected electrically to it, was also measured to ensure that no blockages or non-linear I-V characteristics were present. To obtain the latter, a voltage ranging from –0.5mV to +0.5mV was applied, which was much greater than any characterisation voltage to be applied later. A typical resistance value obtained was between 6kΩ and 10kΩ. Note that this value is an order of magnitude greater than that measured, under similar circumstances, after the ohmic step was completed. The now constrained 2DEG produces a narrower

and hence, more resistive path for the electrons than was measured for the unconstrained 2DEG.

3.5 Scribing, Mounting and Bonding

After evaporation of the gates, the 2-teminal checking procedure was repeated for a selection of Hallbars, to ensure that the deposition of the gates had not affected their characteristics. The Hall bars chosen for electrical testing were, of course, those that appeared of higher quality with no areas missing, etc. The aim of this extensive testing was to obtain a uniform, eight-terminal, measuring Hallbar, whose characteristics are well known at room temperature. This was to minimise the emergence of unusual characteristics in the measurements under cryogenic conditions. The time, expense and frustration of rectifying these during cryogenic work were worth the extra time and effort expended in checking at room temperature.

Scribing and cleaving cut out the selected 'device' chip, which contained three Hallbars. Two of the Hallbars had different gate lengths (100 and 600nm) and the other had none. This was done to confirm that the mere presence of the gate, and hence its Schottky contact, had any effect on the underlying characteristics. Latterly, three Hallbars, with different orientations, but the same standard gate length, were also fabricated in 'device' chips. Possible anisotropy in transport properties could be checked in this way.

The chosen chip was then thoroughly rinsed with a flow of acetone in the cleanroom. This was vitally important, as scribing dust could lodge between the ohmic pads and the eventual wire bonds (see

later). This could degrade both the physical stability and the electrical properties of the bonds. The electrical junction could acquire a large capacitive component, which would only be apparent when the more sophisticated, phase sensitive equipment was applied during the later cryogenic measurement stage (see chapter 4).

The cleaned chip was then glued into a chip carrier that offered up to 18 insulated gold bonding pads. The carrier is designed to permit convenient external electrical connection to the ohmic pads see fig.3.5.1.

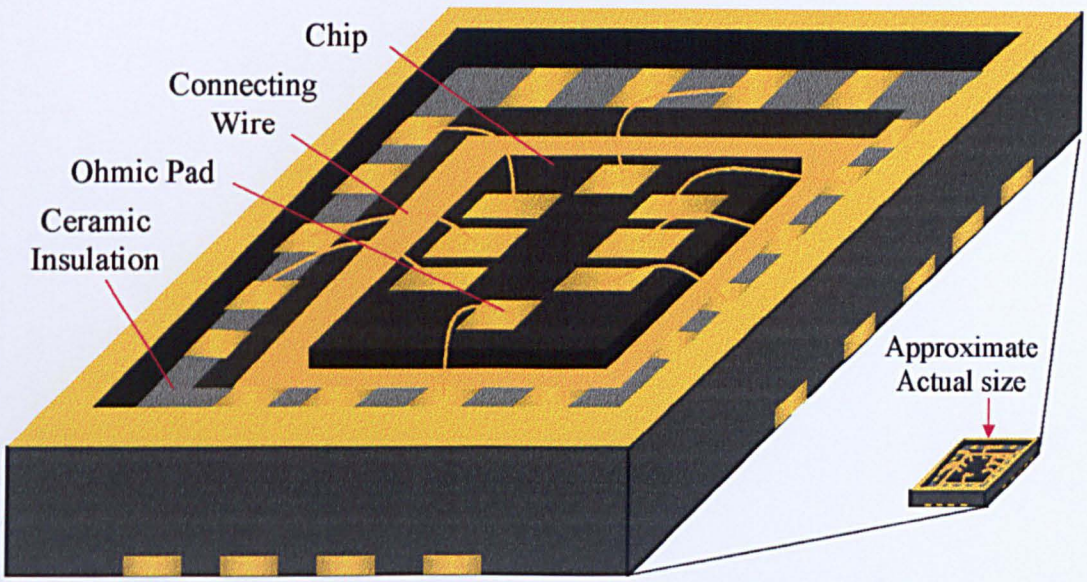


Figure 3.5.1 The final scribed, glued and bonded chip in its chip carrier housing. This is the unit loaded onto a sample rod, which in turn, is inserted into the cryogenic, or room temperature environment system.

The glue (Bostic®) was thinned with a flow of acetone to ensure a thin and flat mounting for future bonding. After allowing several hours for the glue to dry, the bonding process could then commence. The bonder used was a Kulicke and Soffa Industries, Model 4123, and permitted, in each cycle, two ultrasound-assisted bonds. The first was made to a gold pad on the chip carrier, then the second to the relevant ohmic pad on the

chip. Whilst the bonding wedge was directed between the two bonding sites, it fed through a length of fine, gold wire. The settings applied for the two bonds were: (table 3.5.1)

Bond	Search	Force	Time	Power
First bond	**10**	**3.5**	**6.0**	**4.0**
Second bond	**10**	**5.0**	**4.0**	**5.7**

Table 3.5.1 Settings applied on the bonder.

Note that the chip ohmics, the site of the second bond, required slightly more force and power in order to make an acceptable bond. The device was now ready for measurement which, is described in chapter four.

3.6 Summary of the Fabrication Process

The following table summarises the complete fabrication process for those readers who are familiar with the principles, and who only require the procedure.

	FABRICATION LEVEL		
Procedure	**OHMIC**	**ISOLATION**	**GATES**
Cleaning	1. Wafer blown with N_2. 2. Successive wafer submersion in beakers of acetone and IPA, each for 5 minutes and assisted by ultrasound water bath. 3. Rinse with IPA and dry with N_2.		

/Over page

Resist Application	4. 15% Aldridge in chlorobenzene, 5000rpm for 60 seconds, baked for 45 minutes at 180°C. 5. 4% Elvacite in xylene, 5000rpm for 60 seconds, baked for 140 minutes at 180°C.		
Electron Beam Printing	6. Transferred *WAM* CAD file to *VAX* environment. 7. Activated *Readfile* file conversion program. 8. Activated *Cats* program and *Cats Writefile* conversion program. *Cats* software settings:- Resolution- 0.1μm; Extent-all; overlaps-yes. 9. Activated *BWL* program. *BWL* software settings:- Dose- 300μCcm^{-2}; Spot size- 160nm; kvolts-50. Isolation and gates contain extra registration info. 10. Transferred (.com, .iwfl) files to beam writer.		
Development	11. MIBK/IPA (1:1), 24°C, 60 seconds. Then as 3.		
Deoxidation	Wafer in 80 ml Concentrated HCl/ 20ml RO water/ 1ml FC93, for 30 seconds.	None	None
Metallisation by Evaporation	Ohmic 7:-Ni-8nm, Ge-120nm, Au-130nm, Ni-80nm, Au-250nm.	None	Gates1:-Ti-33nm, Au-160nm.
Lift-off	In acetone, at 45°C, for 1hour	None	In acetone, at 45°C, for 1hour
Etching	None	Wafer in 100ml RO water/ 1ml NH_3 solution/ 1ml 30% H_2O_2/ 0.7ml FC93.	None

3.7 References

1. Cho A.Y., Applied Physics Letters, **19**, 467, (1971).
2. Cho A.Y., Arthur J.R., Progress in Solid State Chemistry; Somorjai G., McCaldin J. (eds.), vol. **10**, 157, Pergamon, (1975).
3. Cho A.Y., Journal of Vacuum Science, **16**, 275, (1979).
4. Ploog K., Crystal Growth, Properties and Applications; Freyhardt H.C. (ed.), vol. **3**, 73, Springer-Verlag, (1980).
5. Stradling R.A., Klipstein P.C. (eds.), Growth and Characterisation of Semiconductors, Adam Hilger, (1990).
6. Pearton S.J., Abernathy C.R., Ren F., Topics in Growth and Device Processing of III-V Semiconductors, World Scientific, (1996).
7. Kelly M.J., Low-Dimensional Semiconductors,- Materials, Physics Technology, Devices, Oxford University Press, (1995).
8. Jenkins T.E., Semiconductor Science, Growth and Characterization Techniques, Prentice Hall International, (1995).
9. Chang C.Y., Kai F., GaAs High Speed Devices,- Physics, Technology, and Circuit Applications, John Wiley, (1994).
10. Tsao J.Y., Materials Fundamentals of Molecular Beam Epitaxy, Academic Press, (1993).
11. Parker E.H.C. (ed.), The Technology and Physics of Molecular Beam Epitaxy, Plenum Press, (1985).
12. Joyce B.A., Kinetic and Surface Aspects of M.B.E., in Molecular Beam Epitaxy and Heterostructures; Chang L.L., Ploog K., (eds.), Martinus Nijhoff, (1985).
13. Petroff P.M., Cho A.Y., Reinhart F.K., Gossard A.C., Wiegmann W., Physical Review Letters, **48**, 17, (1982).

14. McElhinney M., Stanley C.R., Electronic Letters, **14**, 1302, (1993).

15. People R., Bean J.C., Applied Physics Letters, **47**, 322, (1985).

16. Matthews J.W., Blakeslee A.E., Journal of Crystal Growth, **27**, 118, (1974).

17. Tedesco C., Zononi E., Canali C., Bigliardi S., Manfredi M., Streit D.C., Anderson W.T., I.E.E.E. Transactions on Electron Devices, **40**, 1211, (1993).

18. Elman B., Koteles E.S., Melman P., Jagannath C., Lee J., Dugger D., Applied Physics Letters, **55**, 1659, (1989).

19. Elman B., Koteles E.S., Melman P., Ostreicher K., Sung C., Journal of Applied Physics, **70**, 2634, (1991).

20. Asenov A., Babiker S., Cameron N., Murad S., Holland M.C., Beaumont S.P., Basics of Pseudomorphic HEMTS Technology And Numerical Simulation, in Devices Based on Low-Dimensional Semiconductor Structures; Balkanski M., (ed.), Kluwer Academic, (1996).

21. Look D.C., Jogai B., Kaspi R., Ebel J.L., Evans K.R., Jones R.L., Nakano K., Sherriff R.E., Stutz C.E., DeSalvo G.C., Ito C., Journal of Applied Physics, **79**, 540, (1996).

22. Kaspi R., Evans K.R., Applied Physics Letters, **67**, 819, (1995).

23. Chattopadhyay K., Aubel J., Sundaram S., Ehret J.E., Kaspi R., Evans K.R., Journal of Applied Physics, **81**, 3601, (1997).

24. Evans K.R., Kaspi R., Ehret J.E., Skowronski M., Jones C.R., Journal of Vacuum Science and Technology B, **13**, 1820, (1995).

25. Toyoshima H., Niwa T., Yamazaki J., Okamoto A., Applied Physics Letters, **63**, 821, (1993).

26. Dehaese O., Wallart X., Mollot F., Applied Physics Letters, **66**, 52, (1995).

27. Gerard J-M., Le Roux G., Applied Physics Letters, **62**, 3452, (1993).

28. Gerard J-M., Applied Physics Letters, **61**, 2096, (1992).

29. Williams R.E., Gallium Arsenide Processing Techniques, Artech House, (1984).

30. Chang C.Y., Chao T.S., in ULSI Technology; Chang C.Y., Sze S.M. (eds.) McGraw-Hill, (1996).

31. Campbell S.A., The Science and Engineering of Microelectronic Fabrication, Oxford University Press, (1996).

32. Lin P.S.D., Gozdz A.S., Journal of Vacuum Science and Technology, B6, **6**, 2290, (1988).

33. Hatzakis M., Solid State Technology, **14**, 79, (1981).

34. Hatzakis M., Journal of Vacuum Science and Technology, **16**, 1979, (1984).

35. Oro J.A., Wolfe J.C., Journal of Applied Physics, **53**, 7379, (1982).

CHAPTER 4

CRYOGENICS AND MEASUREMENT

4.1 Introduction

A study of the various elastic and inelastic scattering factors affecting electron transport in these p-HEMT samples required that the material be subjected to a wide range of thermal environments. In addition, this must be complemented by the ability to extract the relevant data from these environments, and present it in a form suitable for analysis. Therefore in this chapter, the acquisition, storage and manipulation of the experimental data, acquired at different temperatures are discussed.

This undertaking involved the mutual interaction of two major types of experimental systems - one generating and controlling the ambient temperature, and the other acquiring, manipulating and storing

the relevant measurements from within that environment. These two types of systems will be, largely, considered separately, concentrating on the basic principles applied and the equipment used. The noise that is generated both internally, and coupled in externally, will be considered, along with some techniques for its reduction. The overview of the data manipulation will include both the software and the mathematical techniques employed to convert the raw measurements into the various forms suitable for presentation.

4.2 Cryogenics

4.2.1 Introduction

To generate required experimental temperatures, two types of system designs were utilised, both involving the use of cryogens. A third, less complicated system was used for measurements at room temperature. The cryogenic systems both necessitated the use of helium (He)- one where He gas was continually circulated in a closed circuit configuration and the other where the He was exhausted eventually to the atmosphere. The latter method also made use of liquid nitrogen, which, again, was exhausted to the atmosphere. The two designs of cryostat will each now be considered.

4.2.2 The Helium Closed-Circuit System

This system consists of a water-cooled He compressor unit, CTI-Cryogenics® Model 8200, which supplies compressed He gas at room temperature to a drive unit or cold head, Model 22. There, the He gas is displaced by the action of a piston and the gas can be expanded, and hence cooled, by the opening of an exhaust valve. This cold gas cools a sample holder, which is housed inside an evacuated (<0.1 Torr), aluminium chamber. There is also a resistive heating coil and a temperature sensor. The expanded He gas is exhausted back to the compressor to allow more compressed He to enter the refrigerator. The process is repeated until the desired temperature is attained.

A base temperature of about 8K was available after about two hours of gas circulation. Higher temperatures between 8K and 70K were reached by opposing the cooling effect with the application of resistively generated thermal energy. A given temperature was controlled and maintained with the use of a Cryogenic Temperature Controller, Model DTC 500-SP, from Lake Shore Cryotronics®. This uses proportional (P), integral (I) and derivative (D) control to stabilise the temperature at a set value. This latter parameter is determined automatically by the comparison of the voltage dropped across the semiconductor diode sensor with a set reference voltage appropriate for that sensor and temperature. For example, if the measured sensor voltage is comparatively too high, then extra power will be applied to the heating coil. This power will be delivered at a given magnitude and for a given time according to the difference between the measured and the set voltages, and the manner in which stabilisation is desired, as determined by the P.I.D. values.

External electrical contact was made through vacuum-sealed couplings emerging from the top of the cylinder, and which offered co-axial connections. An additional copper shroud was placed internally around the sample and associated wiring to provide additional thermal shielding from room temperature radiation. Provided an extensive temperature range was not required, either at high or low temperatures, this method was a compact and relatively cheap procedure when compared with the open-exhaust He system to be described next.

4.2.3 The Helium Open-Exhaust System

4.2.3.1 Introduction

The closed-circuit system was only used until a more accurate and wide ranging temperature system became available. This configuration was based on an Oxford Instruments® He^4 cryostat (henceforth referred to as the 'cryostat'), that offers a wide range of temperatures, spanning over 300K, with a base of 1.5K. When compared with the closed-circuit system, it was more expensive to run. However, the straightforward automatic control and maintenance of a temperature (at least 3 significant figures), made this system an attractive alternative. In addition, an integral superconducting magnet offered fields ranging between 0 and 10T. Another similar system, offered milli-Kelvin temperatures through the use of a He^3/He^4 dilution refrigerator, and a 13T maximum field was also available. It was used for some earlier experimental work, but was perhaps over-specified for this work. Subtle

quantum effects, normally studied in this system, would be strongly masked by the intrinsic scattering in the layers considered in this work.

4.2.3.2 The Cryostat

The outward appearance of the cryostat was that of a six-foot high by two-foot wide non-magnetic, stainless steel cylindrical shell. The main features contributing to its performance include an outer layer of super-insulating foam, surrounding concentric, outer liquid nitrogen (N_2), and inner helium (He^4) baths, combined with layers that are evacuated to reduce conduction and convection. At the centre is the variable temperature insert (VTI), see also below, which accommodates a sample rod, which supports the bonded sample. Also housed on this rod are a temperature sensor, connective wiring and a switching box.

Concentrically surrounding the lower part of the VTI, and hence the sample, was the superconducting magnet. It consists of a number of concentric solenoids composed of superconducting niobium-titanium filaments within a stabilising copper matrix. The higher field magnet uses inner coils of niobium-tin, which has a higher critical field than niobium-titanium. The net magnetic field produced is very uniform over the sample space, with its direction lying parallel to the cylindrical axis of the sample space. The field was controlled by an Oxford Instruments®' Power Supply, Model PS120-10. This can energise the magnets by delivering up to 120 amps at a voltage of up to +10V and de-energise by absorbing the energy at up to –10V. The sweeping rate can be adjusted to suit experimental conditions and, in this work, was always set at 0.3T per minute. The smallest step size is around 1mT. The

chip carrier was mounted into the rod such that the plane of the Hall bars was perpendicular to the magnetic field direction.

The V.T.I. and the Cooling Procedure

Apart from housing the sample rod, the VTI, as its name suggests, can provide the sample with a wide range of temperature environments. Once the sample rod was inserted, and the inner sample space evacuated, the sample was allowed to thermally stabilise for about twenty minutes to the ambient temperature. Cooling too rapidly can induce undesirable effects on the apparent mobility of the charge carriers within the device. To lower the temperature from this mark required the input of helium from its bath into the sample space, via an adjustable needle valve. Just sufficient for the sample space pump to act upon, was enough to initiate a slow cool - with the local pressure increasing only slightly. As cooling progressed, more helium had to be pumped through to maintain a cooling rate of about 150K per hour. As the vapour pressure of the helium in the sample space is still generally very low with respect to that of the atmosphere, the base temperature does not limit at 4.2K, but continues downward. However, to attain the base temperature of about 1.5K, a lower helium pressure was required than used during cooling. Constricting the needle valve, whilst maintaining the exhaust rate, lowered the vapour pressure to the required level. Provided this rate of input of helium, and its pumping was maintained, then the sample temperature remained approximately constant.

Variable Temperature Control

Any temperature above the base was accessed and maintained by the use of an Oxford Instruments®, Intelligent Temperature Controller Model 502 (ITC 502). In response to the input parameters of the set point temperature and the correspondingly appropriate PID values, the controller will automatically adjust the λ-point needle valve and/ or supply resistive heat energy at the required rates to achieve and maintain the desired temperature. In practice, the needle valve opening was set manually via the ITC 502 Controller, because a zero offset existed, and only the heater power was controlled automatically. The system controlled in response to changes in a temperature sensor fixed to the wall of the sample space near the helium inlet line rather than to the one placed near the sample. Provided the temperature difference between the two sensors was not too great, this was a minor problem. However, at higher temperatures, the temperature as signified by the sensor of the sample space was, at times, up to 100K lower than that measured near the sample. As a result of the large settling time constant, the target temperature took many hours to attain and stabilise. Low target temperatures were less affected by this discrepancy, as by the time such levels were attained, the two sensors were approximately in thermal equilibrium. However, by switching the heater system to manual temporarily, one could reduce greatly this delay.

The Sample Rod

The bonded chip carrier, see fig. 3.5.1, was placed into a depression, set into a block of Tufnol. This socket contained the 18

possible electrical contacts, positioned to match the 18 contacts on the chip carrier. Also found here was a spring-loaded copper thermal earthing pin, which contacted with the base of the chip carrier, as did the electrical contacts when a screw-mounted pressure pad was brought to bear on the top rim of the chip carrier. The electrical wires from the 18 contacts wended back along the hollow rod and terminated at a switch box, which permitted either connection to the external measuring circuitry or electrical earthing to the cryostat, see sect 4.3. The wires emanating from the sample temperature sensor also took this path, but exited at a different port, finally terminating at the ITC 502 or a multimeter if direct resistance values are wanted.

4.3 Measurement

4.3.1 Introduction

Should the reader wish only to consult the relevant circuits employed, without the digression of equipment descriptions, those details are confined to section 4.3.5. However, the inter-relation of the various systems is shown schematically in figure 4.3.1, which is referred to in section 4.3.2.1.

In order to access, measure, manipulate, and store the relevant data arising from within the cryogenic systems and elsewhere, two basic system combinations were used. They are:

1 Two- or four-terminal resistance or conduction measurements emanating from the cryogenic equipment.

2 Four- terminal resistance measurements arising at room temperature.

The measuring equipment, along with the controlling computer, was housed on a portable trolley, with the IEEE interconnections permanently made. In this configuration, the trolley could be moved to the site of the cryogenic or other equipment, which is much less portable or immovable. The power supplies for the production of the magnetic fields were permanently sited close to their magnets. Re-calibration would be required if each magnet did not have a dedicated power supply.

4.3.2 Instrumentation for Cryogenic Measurement

4.3.2.1 Introduction

Most measurements were made in the cryogenic environment, with the room temperature acquisition much less common. The systems consisted of:

1 Two E, G & G® Princeton Applied Research, Model 5210 Lock-in Amplifiers.

2 One Hewlett Packard®, Model 3245A Universal Source.

3 Two E G & G® Parc, Model 113 Pre-amplifiers.

4 One Keithley®, Model 236 Source-Measure Unit.

5 One I.B.M.®- compatible P.C.

Either Labview® or Turbo-Pascal® software controlled this equipment, with the former offering a graphical user interface (GUI). Some simple additional external circuitry was also incorporated in the systems and will be described below, in section 4.3.5. The manner in which the various sub-systems interact and connect to the available equipment and circuitry is shown in fig. 4.3.1.

There were basically four inter-relating sub-systems accessing the four available temperature environments:

1 The most sophisticated system or route invoked was that where the oscillator signal was directed through low-pass, signal filter circuitry, then potentially divided, before being passed into the sample which was contained within a cryostat - itself housed in a shielded room. The cryostat inside contained the 13T magnet and the dilution refrigerator when required. The gate voltage was directed through an L.C.R. (inductance-capacitance-resistance) filter, see fig. 4.3.5a. The measured signals were directed back through the signal filter after being pre-amplified.

2 Much more convenient for this work, was the second system. It used a similar cryostat (but not in a screened room), with less complicated C.R. gate filtering, see fig.4.3.5b. However, at the signal levels appropriate for these measurements, the performance was as good as in 1 above.

3 The closed-circuit system was similar to 2 above, but with no pre-amplification of the measured signals.

4 Lastly, the room temperature system was similar to 3 above.

An account of the components of these systems will now follow.

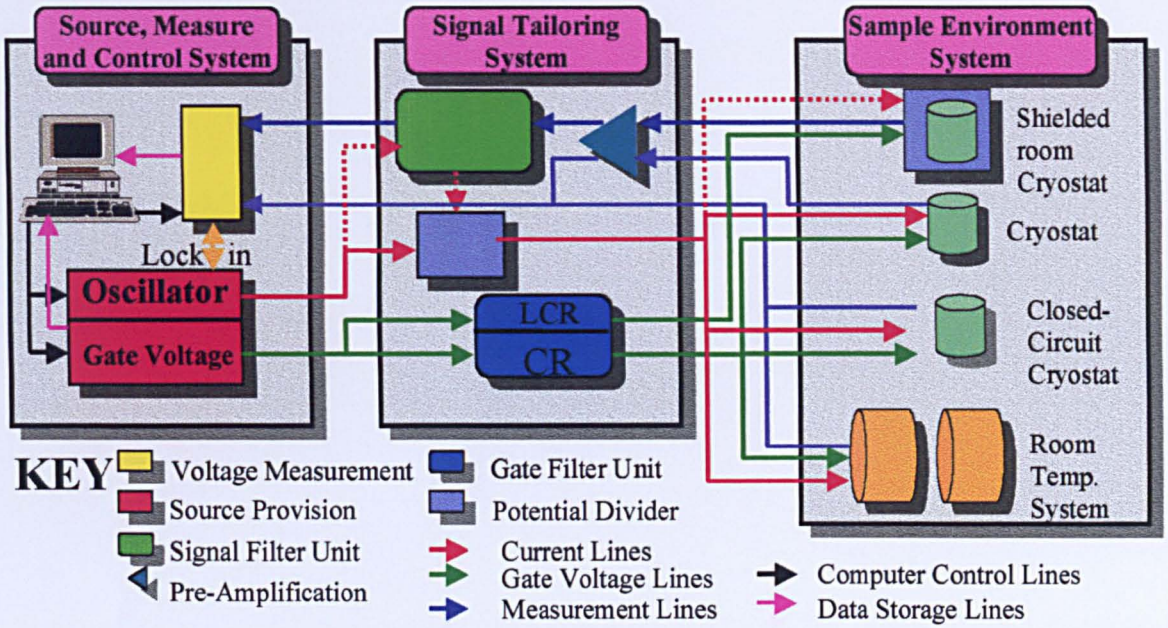


Figure 4.3.1 The inter-relationship of the various sub-systems used to acquire data in this project. Note that some routes through the components are more complicated than others, but, provided care was taken with the procedure, all system routes performed to a similar level. A more detailed explanation can be found in the main text.

4.3.2.2 Lock-in Amplification and Measurement

The two lock-in amplifiers had a dual purpose. The first was to provide the channel layer of the sample with a low frequency a.c. excitation current. It was directed between two designated ohmic contacts, usually the source and drain, as shown in fig. 3.2.2(a or c). The second purpose was that of measuring the voltage dropped across any two designated ohmic contacts and across an external resistor. If the two ohmic contacts were the same as those receiving the current, then a '2-terminal measurement' would be made. Two different contact pairs would entail a '4-terminal measurement'.

The lock-in amplifiers are capable of providing an a.c. signal of voltage 0V-1.9V, and frequency 0.5Hz- 120kHz. However, for all the measurements acquired in this work, only a 1V, 18Hz signal was used. They are also capable of measuring voltages within a wide range of full-scale deflection sensitivities. To exploit this, the Labview software includes an acquisition pause to allow sensitivity range changes to be made and hence, maximise the accuracy. This was important, as conductance levels could become very low, particularly in the strong localisation regime. The lock-in amplifiers use an internal phase-sensitive detector (P.S.D.), which detects the measured signal at a reference frequency, also 18Hz and in phase with the oscillator. Its output contains a contribution representing the signal of interest, but also a.c. components representative of the noise and interference. Low-pass filters can then reduce the magnitude of these latter components, improving the signal to noise ratio. As expected, the measured signal always contained a small out-of-phase component because reactive contributions existed, in the measurement circuitry. However, good

circuit design and careful fabrication, kept these magnitudes to a minimum. For example, failure to remove the scribing dust from the surface of the ohmic pads before bonding can add a strong capacitive component to the voltage measurements. Generally, with good practice, they never exceeded about 5% of the main in-phase components at high carrier densities, and were ignored as far as acquisition was concerned.

4.3.2.3 The Universal Source

A variable (or single-valued) gate voltage was supplied by the Universal Source, which delivers a very accurate and stable d.c. level. The Labview software can specify the voltage range applied, and the increment or decrement step size and time to cover that range. It also ensures that the acquisition of the measured voltage only occurred at the end of a significant step interval to allow for the output to settle in response to the new gate voltage. A typical range applied was +0.5V to –3.0V, decrementing in steps of either 0.02V or 0.05V. A settling time of ten seconds was usually just sufficient. There was a longer settling time required when a change of sensitivity range was made. This, however, was not of concern, as the pause function was available. If required, the Universal Source could also supply a constant (gate) voltage, whilst some other parameter was varied, for example, the magnetic field.

Any d.c. voltage signal applied underwent filtering as close to the cryostat as possible. The resistance-capacitance (RC) or resistance-capacitance-inductance (RCL) circuits filtered out the externally coupled, higher frequency noise. These circuits are shown in figs.

4.3.5(a and b) in section 4.3.5. The series resistance component (R) also ensured that the current existing in that line was kept low, should the energy barrier of the Schottky contact insufficiently prevent a transverse current flow through the sample layers. On occasion, this occurred when a large gate voltage was applied, but damage to the gate was prevented owing to the limit on the gate-leakage cuurent.

4.3.2.4 The Source Measure Unit

The Keithley, Model 236 Source Measure unit was used only on occasion to implement two measurements:

1 A determination of the gate-leakage current, resulting from the applied gate-channel voltage range, when it was suspected to be significant.

2 The provision of a d.c., rather than the usual a.c., excitation current for studies in the strong localisation regime.

The former measurement was controlled by a Turbo-Pascal software routine, whilst the latter was directed from within the Labview suite.

4.3.3 Room Temperature Measurements

Some measurements at room temperature were initiated to establish the effect on mobility of the addition of inelastic scattering due to phonons, and to provide some useful information for device engineers. Resistance measurements were restricted to those derived from the application of a low and constant magnetic field to the sample.

In other words, only the Hall mobility was available from such resistance measurements. These were calculated using the simple circuit principles to be described later in section 4.3.5.

A Newport Instruments Electromagnet type E, between whose poles the sample could be placed, provided the field. The sample was bonded, as usual, in a chip carrier, see fig. 3.5.1, which, in turn, was mounted in a hinged 18-pin plastic chip carrier holder. This holder was placed co-axially between the poles of the magnet. Its pins were soldered to a section of a breadboard, from which connection was made to a more remote switch box via a hollow metal tube. The switch box was clamped to a framework above the magnet and supported BNC sockets, permitting co-axial electrical connection to the measuring equipment. The overall system used is shown as part of fig. 4.3.1, and was much the same as that used by others for acquiring voltage measurements under cryogenic conditions. The main elements missing were the signal filters and the pre-amplifiers- both found unnecessary when the electromagnet was carefully sited remote from local noise sources. The magnet was powered by a constant current supply delivering, in this case, 5.00A, which provided a uniform field of 0.70±0.01T.

4.3.4 Noise Reduction Measures

If the measured voltages have unacceptably high noise components superimposed, then any characteristic parameter, such as mobility, derived from this data, will be unacceptably 'noisy'. In this work, the reduction of noise was not as critical as it would be for some

other cryogenic measurements involving quantum effects, but basic steps were still required.

All electrical connections external to the sample housings were co-axial and twisted where possible. In order to minimise electron heating or inter-subband scattering within the sample channel's 2DEG, the applied currents must be limited. In consequence, the measured output voltages were also small and, as such, could be overwhelmed by noise infiltration, particularly if the leads carrying these small signals were excessively long.

Electron heating and its resulting noise can also be induced by excessive electromagnetic (e.m.) signals coupling externally to wires at any point in the input half of the circuit. Therefore, the wire connections were kept as short as possible after potential dividing and before the returning signal was pre-amplified. This latter action boosted the signal (1000X), and, in combination with the low-pass signal filter box, improved the signal to noise ratio. As stated earlier, one of the cryostats was housed in an electromagnetically shielding room, which helps to prevent r.f. noise coupling to the wires carrying the small signals, but it made little difference in this work.

Although less critical, the wires carrying the amplified measured signals from the pre-amplifiers back to the voltmeters (lock-in amplifiers) were twisted and also not unnecessarily long. The oppositely directed wire, carrying the full (1V) oscillator input between its lock-in output and the potential divider, was also kept as short as possible. These connections were never housed within the r.f. shielded room. Locally produced electromagnetic noise spikes were particularly destructive, thus any measure that avoided or reduced their effect had to be implemented. However, as implied above, the pre-amplifiers were

not available for transportation on the trolley. In consequence, to avoid locally produced light switching and other electrical noise sources, the siting and timing of the room temperature measurements, in particular, was critical to their success. Fluorescent lighting itself was not an apparent problem- only its switching on and off.

Damping the cryostat pump lines with lead foam, and placing the pump and the cryostat on rubber matting reduced mechanical vibrations that emanated from the pumps. Finally, the cryostat within the shielded room was earthed to this shielding, whilst the other cryostat was connected to the common earth of its magnetic power supply. The whole electrical circuit was designed to have the same earth point. This reduced the opportunity of earth loops developing where a spatial distribution of slightly differing earth potentials arose.

4.3.5 The Measurement Circuits

The circuit details are confined to this section should the reader only wish to consult these without the digression of equipment descriptions or noise suppression techniques. The main elements of the external circuitry are shown schematically in fig. 4.3.2. For all of the a.c. measurements in this work, an oscillator applied a 1.000V, 18Hz signal. This level of signal was much too high to pass through the device, and would effect significant electron heating and inter-subband scattering. This high level was maintained for most of its passage to the sample

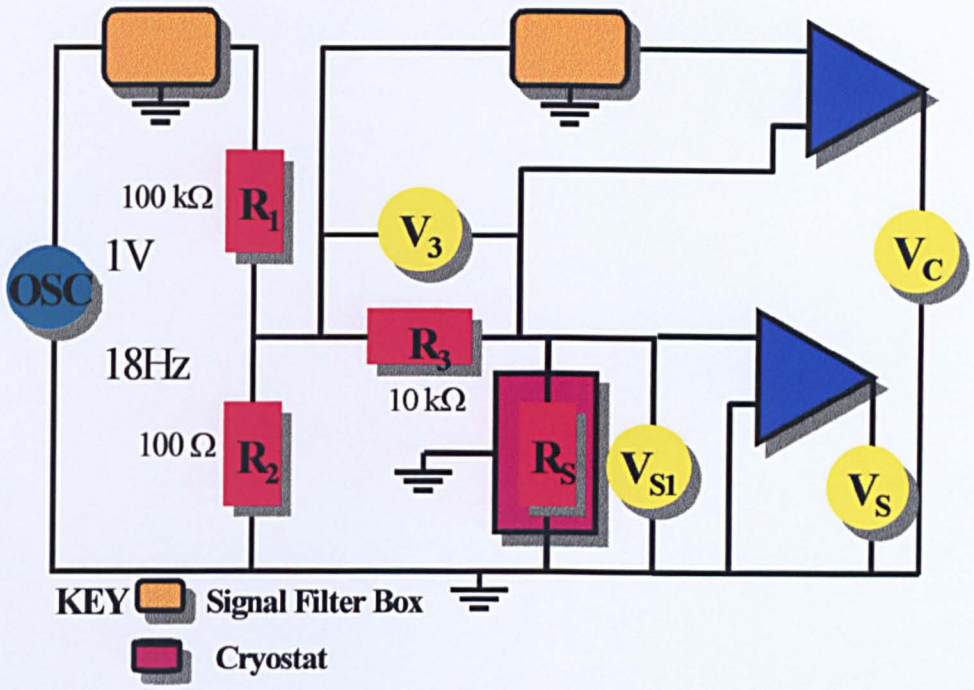


Figure 4.3.2 A schematic illustration of the external circuitry, showing the potential divider (R_1, R_2), the series resistance (R_3), the sample (R_S), pre-amplifiers and signal filtering. The 'optional' components are shown with solid-line borders. The complimentary gate voltage sub-system is shown in fig. 4.3.4.

in order to keep the signal to noise ratio relatively high, see section 4.3.4. However, it must be significantly reduced just before entering the sample. A potential divider, composed of two variable resistors, R_1 and R_2, achieved this reduction, outputting a 1mV, 18 Hz signal, see fig. 4.3.2. The two resistance values applied were 100kΩ and 100Ω respectively.

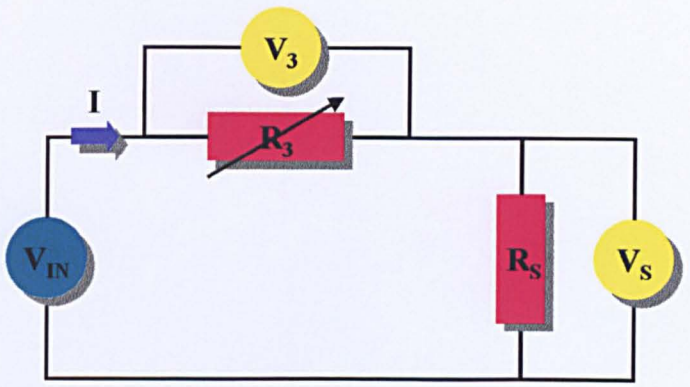
An adjacent variable resistor, R_3 that was always set to 10kΩ, was placed in series to further reduce the current, which was then passed into

the sample, denoted by R_S. The voltage signal, V_{S1}, dropped across the sample, was sometimes boosted by a factor of 1000X by a pre-amplifier sited as close to the sample as possible. The boost was effected to improve the signal to noise ratio as described in the previous paragraph, and in section 4.3.4. The amplified signal, V_S, was fed back to the lock-in amplifier that originally provided the oscillator signal, but can also measure incoming voltages. For greater sophistication, the source and measured signals were passed through a signal filter box. Fig.4.3.1 displays the appropriate route, and hence, the system required. In fig.4.3.2, the optional components are denoted with solid-line borders.

The series resistor, R_3, also provided a means of current measurement that is derived from the voltage dropped across it. This voltage signal, V_3, extracted from R_3 (10kΩ) could also be boosted and fed straight back to the other lock-in amplifier/ voltmeter via the signal filter. A typical boosted signal, V_C, measured 4-terminally by this voltmeter, was about 750mV, implying that the voltage dropped across R_3 was 750μV. Therefore, the current through it was 750μV/10kΩ, or about 75nA.

In the room temperature system, the pre-amplifiers were missing and hence, V_3 and V_{S1} were kept small when compared to their levels derived from the systems using the amplification. This drop in measured signal required a commensurate 1000 times rise in the voltmeters' sensitivity, but, provided care was taken with the data acquisition, did not noticeably degrade its quality.

Armed with the two voltages V_C and V_S, the software converts these measurements to a sample resistance, which was conveniently plotted out in real time versus the independent variable, and stored for future use. The simple algorithm used is illustrated in fig. 4.3.3



$$I = V_3/R_3 = V_S/R_S$$

$$\Rightarrow R_S = R_3 V_S/V_3$$

Figure 4.3.3 The 'local' circuit from which the voltage measurements are acquired, and the simple algorithm employed to calculate the sample resistance. V_3 can be regarded as V_C here.

The software calculated the sample resistance by dividing the sample voltage, V_S, by the series resistance voltage, V_C, and multiplying by the selected value of the series resistance.

The sub-system providing the gate voltage is displayed schematically in fig. 4.3.4. The d.c. voltage, derived from the Universal Source supply, was fed initially through a gate filter box, which limits possible damaging currents should gate leakage exist. The voltage applied to the gate ohmics and hence the gate was referred to the source contact. The two gate filter circuits used are shown in fig.4.3.5(a and b). The version using the inductance (a) was integral to the cryostat system housed in the shielded room, whilst the other (b) formed part of the portable trolley system, and hence, all the other systems.

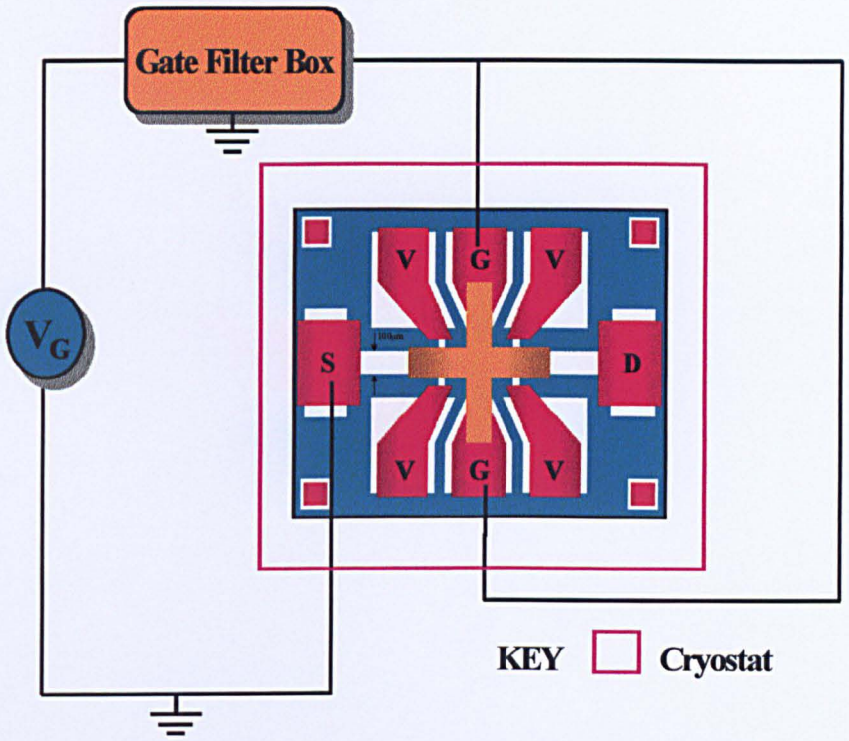


Figure 4.3.4 The schematic gate voltage circuit. Note that the gate voltage is referred to the source of the sample. The cryostat is, of course, optional.

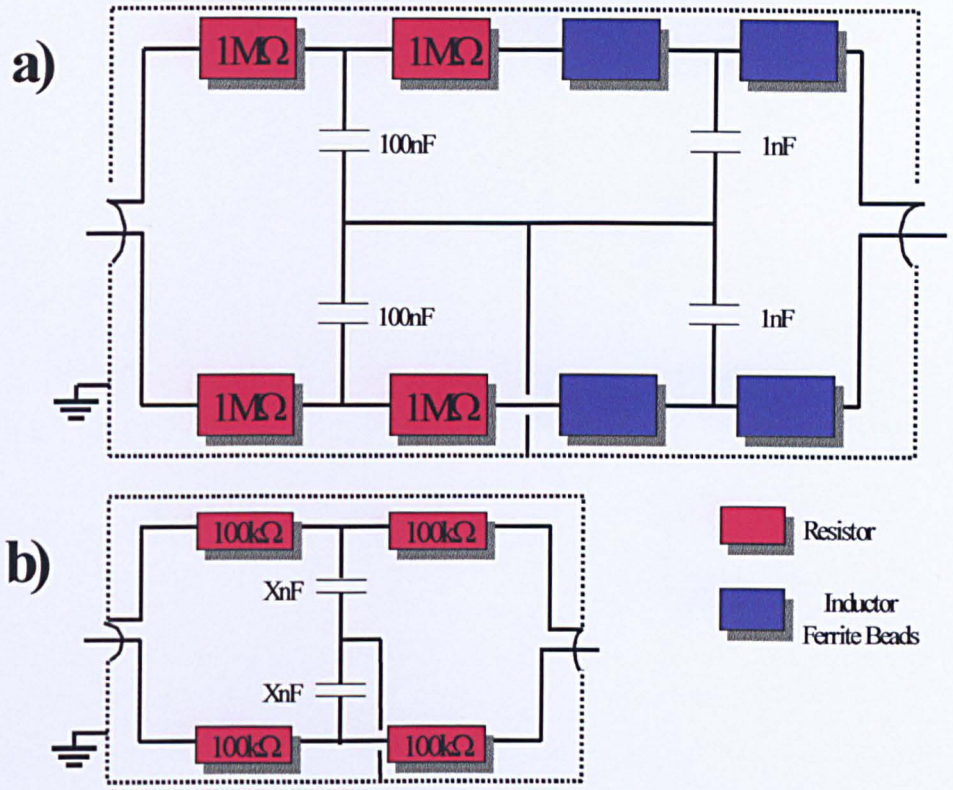


Figure 4.3.5(a and b) The two gate filter circuits used. **a)**- Is the L.C.R. circuit incorporated as part of the system in which the cryostat is shielded. **b)**- Is the simpler R.C. circuit used in all of the other systems. See also fig.4.1. The dashed line represents the metal housing, which is earthed. 'Mirror' circuitry is used as the input and output lines are co-axial. The values (X) of the capacitors used in *b* are unknown.

CHAPTER 5

RESULTS AND DISCUSSION

A The Metallic Regime

5.1 Introduction

The experimental data was acquired at various temperatures from 1.5K to 300K mainly using the classical Hall effect. Measurements made in this way are accurate when only one conducting 'pathway' is significantly activated within the sample layers. This pathway should ideally consist only of electrons in the lowest subband of the 2DEG; however other channels of much lower mobility may also contribute. These can include electrons found in the higher energy subbands of the channel and/or the shallow states of the doping layer. The Hall mobility, calculated from the measured resistance

values, is a weighted-average of all the conducting states and as such, would be misleading if only the mobility of the first subband was sought.

Magneto-resistance or Shubnikov-de Haas (SdH) data acquired at the same time as the Hall data confirmed that no second subband occupation in the channel, or parallel conduction in the doping layer was present in any sample under any cooling conditions at the base temperature of 1.5K. Higher subband occupancy would be revealed by the presence of beats, and parallel conduction by a rising (B^2) background to the plot.

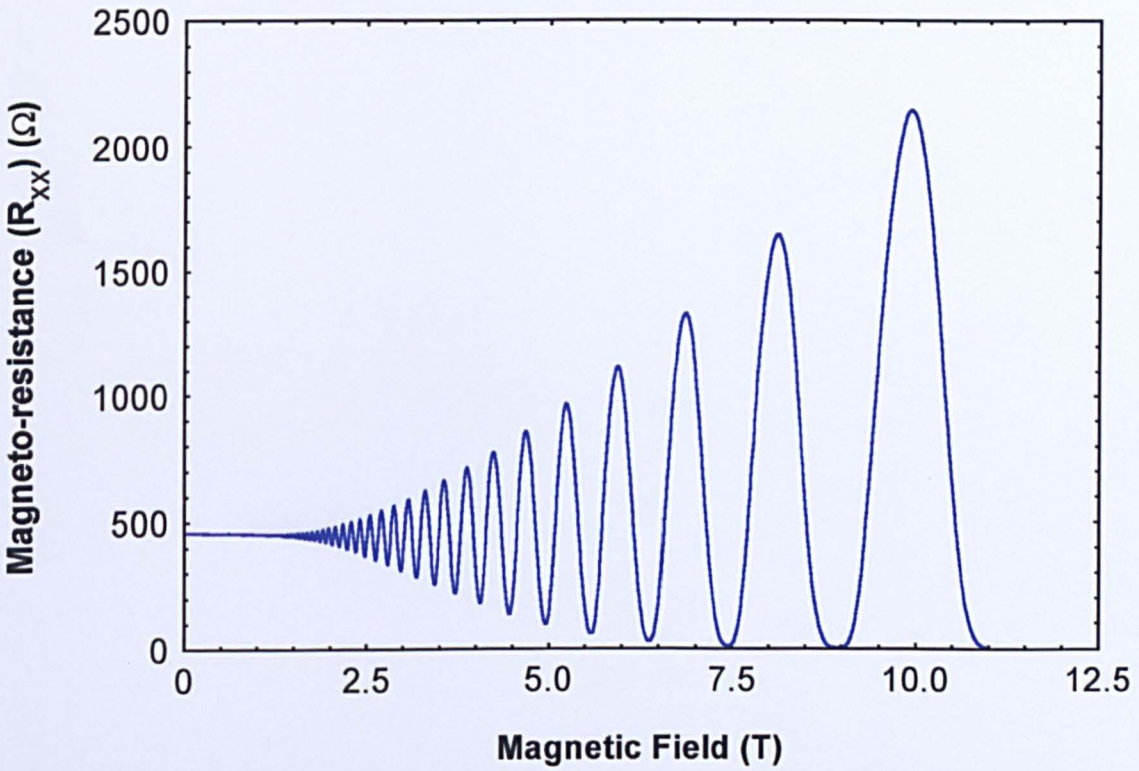


Figure 5.1.1 A typical Shubnikov-de Haas plot acquired for layers of this type. This particular example was of A1047 – a sample with 30% indium in the channel and a 5nm spacer. The data was taken after a normal cool at 1.5K with a zero gate bias applied during measurement. Note the absence of both beats and a rising (B^2) envelope in the magneto-resistance data.

A typical example of an SdH plot is shown in fig. 5.1.1 for a sample (A1047) with 30% channel indium content and a 5nm spacer. If a second

subband was occupied, the fast Fourier transform (FFT) of this SdH data would display a second peak representative of this second population. Only one peak per FFT plot was ever detected, with a typical example (also for A1047) shown in fig. 5.1.2.

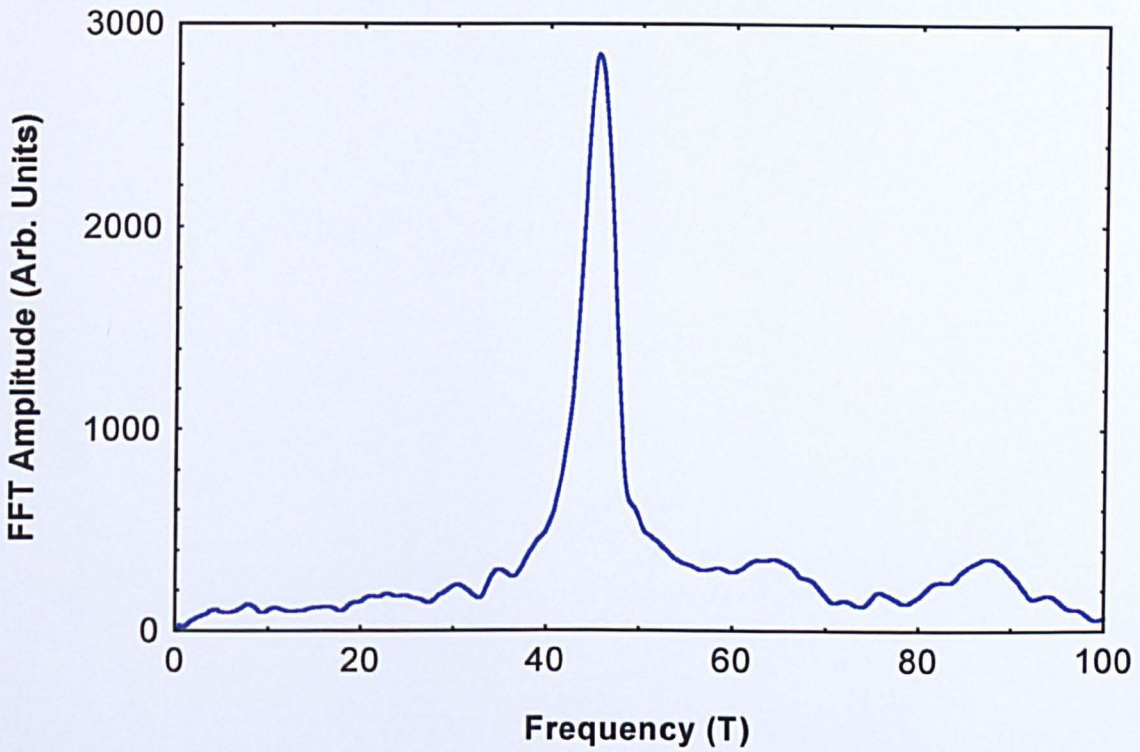


Figure 5.1.2 The fast Fourier transform of the magneto-resistance data in figure 5.1.1. Note the absence of secondary peaks – the hump at about 90T is a harmonic of the major peak.

The frequency at which the peak is found is representative of the first subband electron population calculated using equation 2.2.13. At higher carrier densities, this value was in agreement with the equivalent Hall carrier density to within about 1% derived under the same conditions. The error associated with the SdH data increases with a decreasing channel population due to the decreasing number of oscillations. Based on the above observations, the Hall measurements were applied with confidence to test

the electrostatic model (section 2.1), and then to make other measurements in the metallic regime. The other major applications of the longitudinal magnetoresistance data accumulated in this regime were to analyse the effect of weak localisation where appropriate, and to determine the quantum mobility.

5.2 The Applicability of the Electrostatic Model

The experimental values and theoretical predictions for the carrier density (n), critical doping density (N_δ), threshold gate voltage (V_T^f) and capacitance per unit area (C^f/A) are compared in table 5.2.1. The

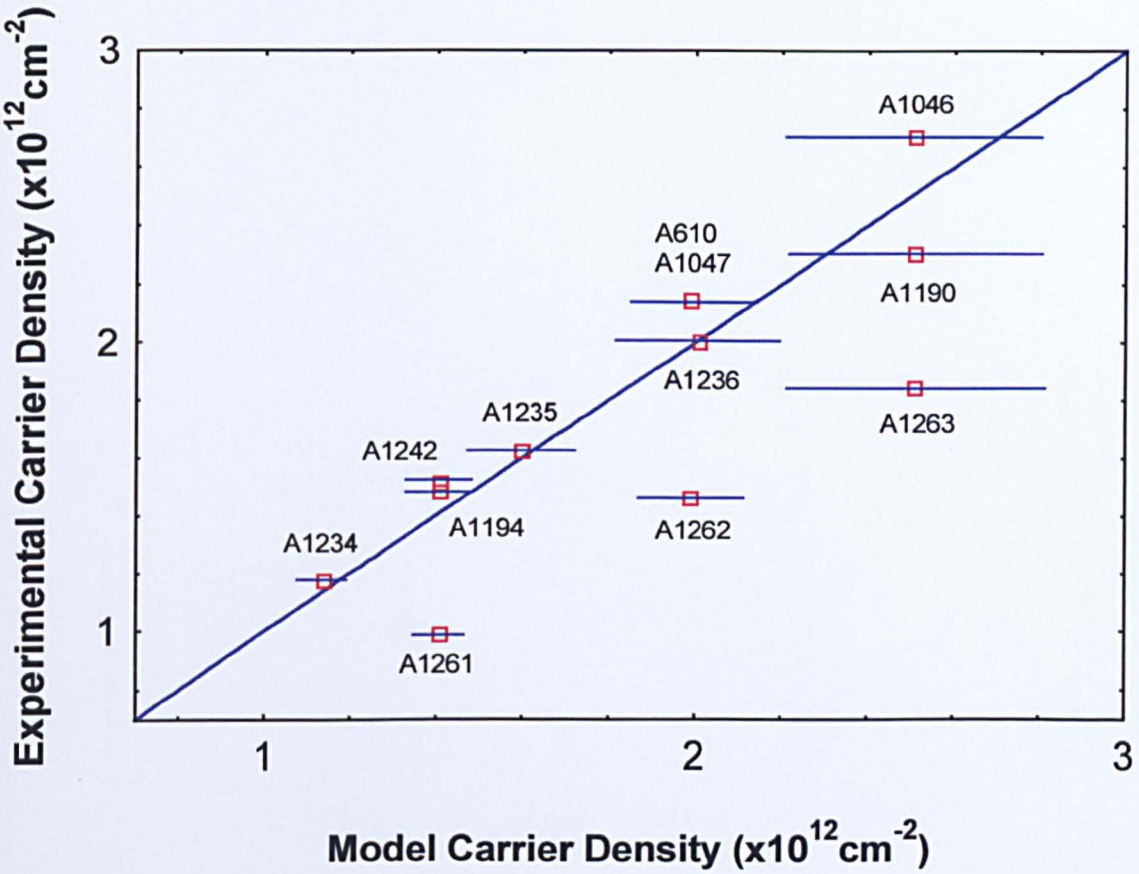


Figure 5.2.1a The relationship between the predictions of the electrostatic model and the experimentally acquired values for the carrier density. The blue line (y=x) represents equality of the two. An effective channel width of 7nm was assumed but 'error' bars are shown assuming this value can range between 6 and 8nm. E_{dd} is 70meV. Samples A1261-3 were deliberately underdoped (see next figure).

theoretical numbers were calculated from equations 2.1.9-12 in section 2.1. Figs.5.2.1(a-d) illustrate more conveniently a direct comparison of

the experimental and theoretical data for all the samples. Fig.1.2.4 in section 1.2.4 shows that the Fermi level does not extend across the full physical width of the channel. Therefore, an estimate of the effective channel width (a) was required. A value of 7nm was chosen from theoretical self-consistent results of the bending of the conduction band-edge in the channel region and its relation to the Fermi energy. The

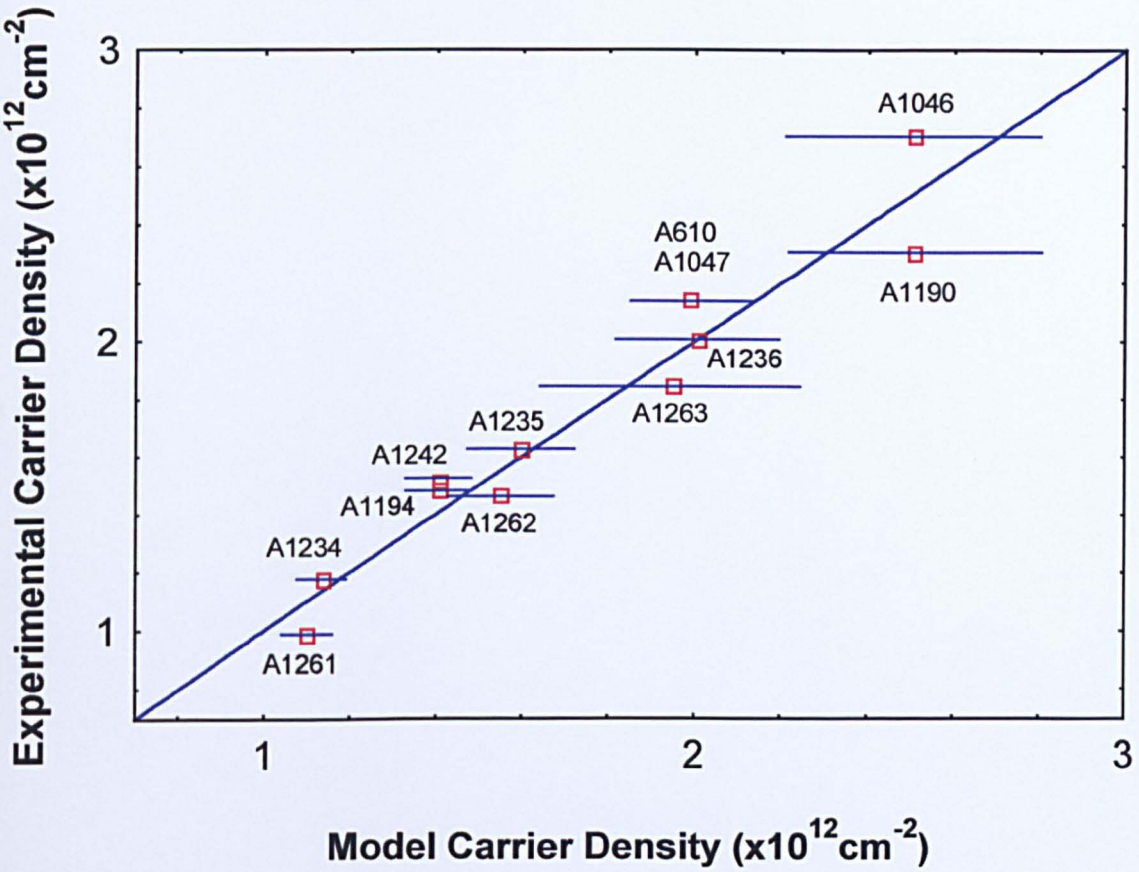


Figure 5.2.1b The same data as plotted in fig. 5.2.1a, but with the three underdoped samples (A1261 (10nm spacer), A1262 (5nm spacer) and A1263 (2.5nm spacer)) assigned a different value for E_{dd} – 150meV (see text). Note how this new value reduces the model density such that the data now lies closer to the y=x line.

error bars relating to the predicted parameters reflects this uncertaintity. The absence of a gate bias during cooling promoted equilibrium occupation of the *DX* centres. As stated in sections 1.4, 2.1 and 2.3, this

occupation also produces donor layer correlation and the pinning of the Fermi level. The latter condition was necessary for the applicability of the electrostatic model. When studying the plots and table 5.2.1, the approximations implemented in both the model and in assigning a constant value for the effective channel width should be remembered.

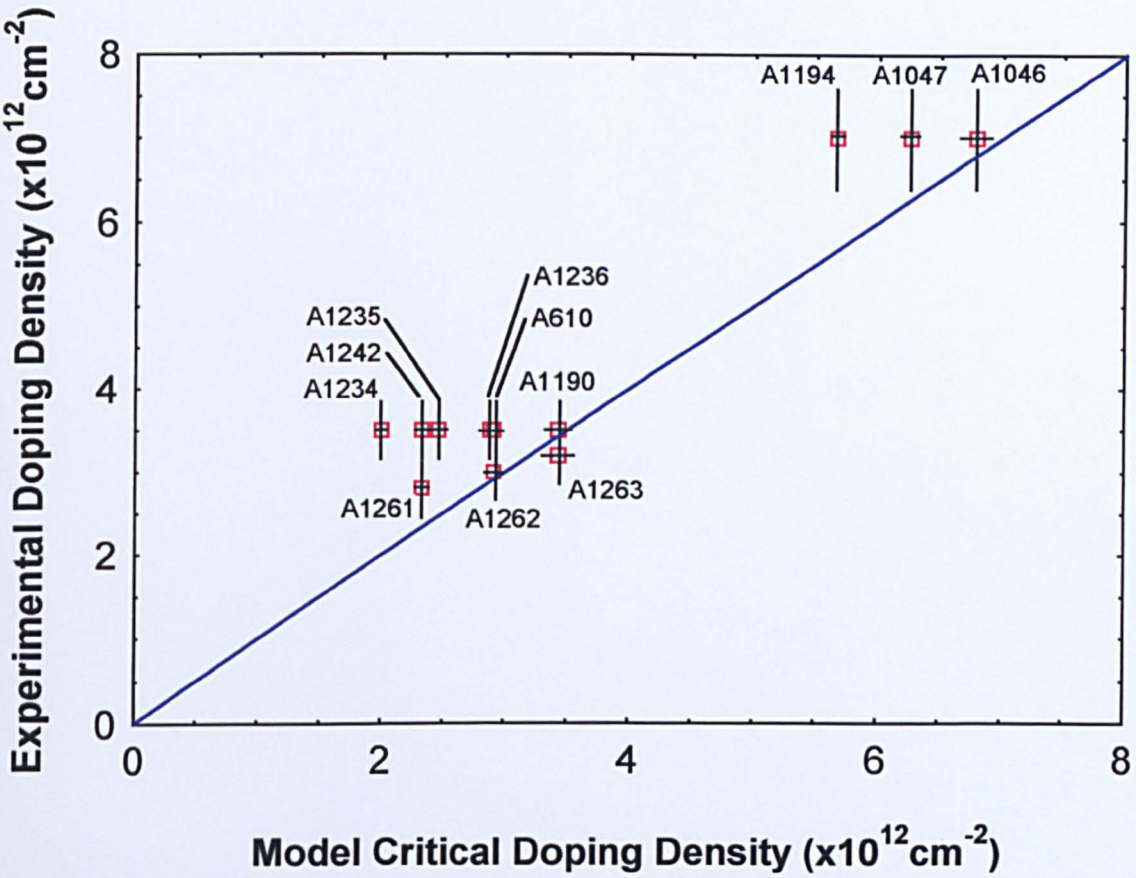


Figure 5.2.1c The relationship between the predictions of the electrostatic model for the critical doping density and the experimental values implemented. The blue line (y=x) represents equality of the two. An effective channel width of 7nm was assumed but model 'error' bars are shown assuming this value can range between 6 and 8nm. Note that most samples are over-doped.

Based on the chosen value (7nm) for the effective channel width (a), excellent agreement between theory and experiment for the carrier density (n_{2D}) was obtained for most samples (fig.5.2.1a). This statement

does not include the samples designated A1261-3 as they were, in effect, deliberately sub-critically doped. All the other samples were analysed assuming a deep-donor energy (E_{dd}) of about 70meV,[1] appropriate for a completely filled lowest *DX* level and a partially filled second level. However, if the deep-donor energy is changed appropriately to about 150meV (the value of the lowest incomplete *DX* level, see section 1.4 and fig. 1.4.1) to accommodate this under-doping, good correspondence is obtained there also (fig. 5.2.1b). The good agreement is not surprising as examination of both the conduction band edge (fig.1.2.4) and equation 2.1.9 reveal. Qualitatively, both GaAs/AlGaAs and InGaAs/AlGaAs layer systems have identical band-edge profiles around the channel-spacer boundary - the difference is only quantitative. (The model was based originally on GaAs/AlGaAs type systems.) Also, from equation 2.1.9, the carrier density is independent of the doping density, hence this will play no part provided both that the pinning requirement is satisfied and that E_{dd} does not change. Any excess donor electrons existing in the doping layer would not affect the electrostatics on which the model was based. The model therefore, should be as accurate for predicting carrier density within pseudomorphic layers notwithstanding any possible effect from the presence of elastic strain.

The threshold gate voltage (V_T^f) and the capacitance per unit area (C^f/A) are less well predicted. As stated in section 2.1, these parameters are closely related as both are connected to the depletion rate of carrier density with gate voltage. The experimentally acquired threshold gate voltage of most samples is more negative than that of the theoretical forecast, (see fig.5.2.1d). The large discrepancy in the gate voltage can be related to the overdoping (fig.5.2.1c) of such samples.

Parameters			n ($\times 10^{12}$ cm^{-2})		N_δ ($\times 10^{12}$ cm^{-2})		V^f_T (V)		C^f/A ($\times 10^{-3}$ Fm^{-2})	
	Indium (%)	Spacer (nm)	Model	Experimental ±1%	Model	Experimental ±10%	Model	Experi-mental ±5%	Model	Experi-mental ±5%
Shallow Channel	30	2.5 A1046	2.51±0.29	2.70	6.79±0.24	7.0	-1.01±0.06	-1.2	3.97±0.13	3.90
		5 A1047	1.99±0.15	2.14	6.26±0.15	7.0	-0.877±0.04	-1.1	3.64±0.11	3.50
		10 A1194	1.41±0.07	1.48	5.66±0.08	7.0	-0.726±0.02	-1.1	3.11±0.08	2.29
Deep Channel	30	2.5 A1190	2.51±0.29	2.30	3.43±0.23	3.5	-3.30±0.28	-3.5	1.22±0.01	1.07
		5 A610	1.99±0.15	2.14	2.91±0.16	3.5	-2.70±0.17	-3.0	1.18±0.01	1.17
		10 A1242	1.41±0.07	1.51	2.33±0.07	3.5	-2.02±0.09	-3.0	1.12±0.01	1.00
	20	2.5 A1236	2.01±0.19	2.00	2.88±0.19	3.5	-2.65±0.22	-4.0	1.22±0.01	1.00
		5 A1235	1.60±0.12	1.62	2.46±0.12	3.5	-2.17±0.14	-3.5	1.18±0.01	0.961
		10 A1234	1.14±0.06	1.17	2.00±0.06	3.5	-1.62±0.07	-4.0	1.12±0.01	0.801
	30 (effectively sub-critically doped)	2.5 A1263	2.51±0.29	1.84	3.43±0.23	3.2	-3.30±0.28	-2.8	1.22±0.01	1.02
		5 A1262	1.99±0.15	1.46	2.91±0.16	3.0	-2.70±0.17	-2.3	1.18±0.01	1.00
		10 A1261	1.41±0.07	0.985	2.33±0.07	2.8	-2.02±0.09	-1.7	1.12±0.01	0.942

Table 5.2.1 Table comparing the predicted parameters of carrier density (n_{2D}), critical doping density (N_δ), threshold gate voltage (V^f_T) and the capacitance per unit area (C^f/A) with those obtained experimentally. See the text for comments. The comparative model and experimental data are also plotted in figs. 5.2.1(a-d).

The predicted critical doping density (N_δ) is useful to monitor whether the actual doping density (N_d) of a given sample is sufficient to ensure pinning of the Fermi level in the *DX* centres and thus validate the model.

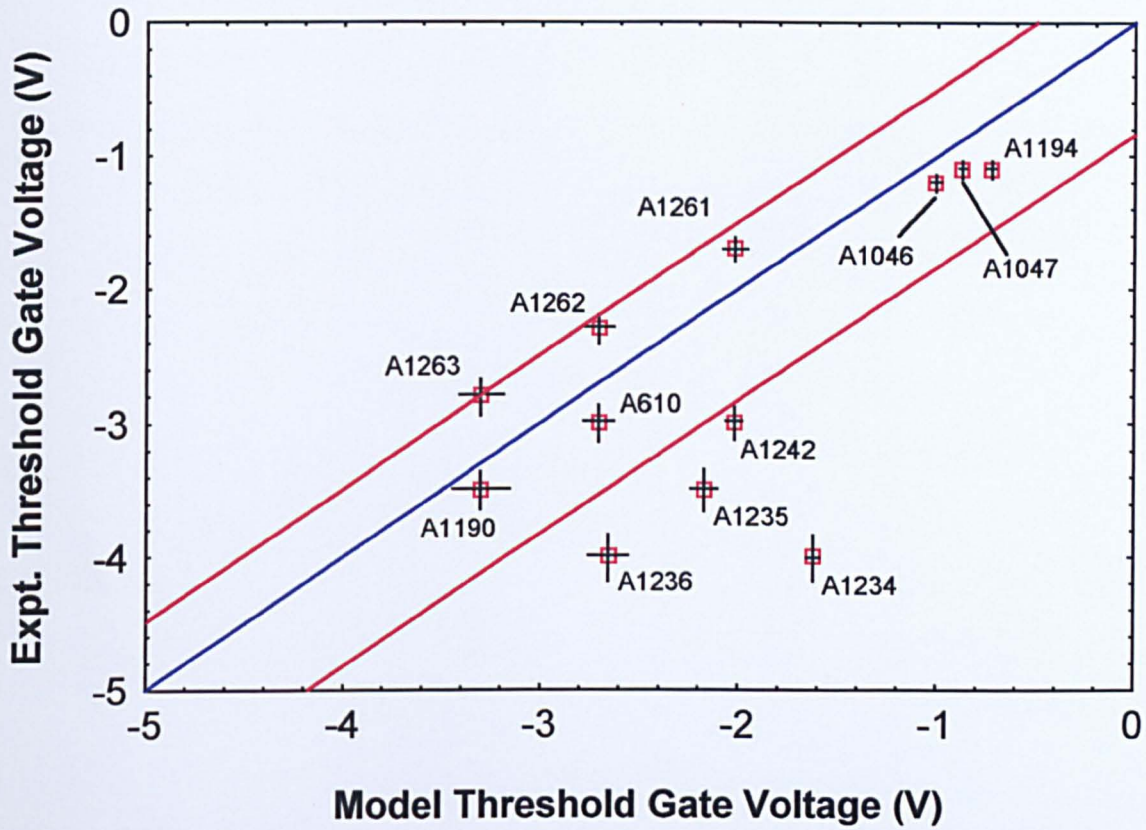


Figure 5.2.1d The relationship between the predictions of the electrostatic model and the experimentally acquired values for the threshold gate voltage. The blue line (y=x) represents equality of the two. An effective channel width of 7nm was assumed but 'error' bars are shown assuming this value can range between 6 and 8nm. Samples A1261-3 were deliberately underdoped effectively. Note that for most samples, the experimental value has a greater magnitude than that predicted. The over-doping of these samples causes this with a resultant shift in the threshold gate voltage. If this factor is removed from the depletion curves, then the data of all the samples lies within the two red lines shown.

From table 5.2.1 and fig.5.2.1c, it is apparent that most of the samples are 'over-doped', and this observation may influence other parameters through the likely over-population of the shallow states of the doping layer. Under

these conditions, the layers would deplete non-linearly when a negative gate voltage was applied. As the gate voltage was made negative, the normal rate of depletion of the channel electrons would be delayed by the initial removal of the loosely bound electrons in the doping layer. These electrons would have to be removed before the gate field could do likewise to the channel electrons. Therefore, the carrier density depletion with negative gate voltage would be initially flattened then curve downwards, developing a linear trend once the extra doping layer electrons were removed. Such features were observed experimentally in some samples, particularly those where the difference between the doping density of the sample and the critical doping density of the model is greatest. As a result, only the linear sections of these plots were used for the calculation of the experimental capacitance. An illustration of this effect is shown in fig.5.2.2 for sample A1242 (blue line).

The claims made above gained further support when samples were bias-cooled, thus increasing the likelihood of electron occupation of the shallow states of the doping layer, when the bias was released. Samples that depleted linearly after a normal cool, developed non-linear characteristics after a bias-cool, with the symptomatic flatter curve at higher density. (Bias cooling effects will be discussed in detail in section 5.4, along with the relevant plots.) Further evidence in favour of this idea was obtained when the sub-critically doped samples (A1261 - A1263) were subjected to a depleting field. All displayed linear depletion over much of the range, as expected for samples in which the doping was near the critical value (fig.5.2.2).

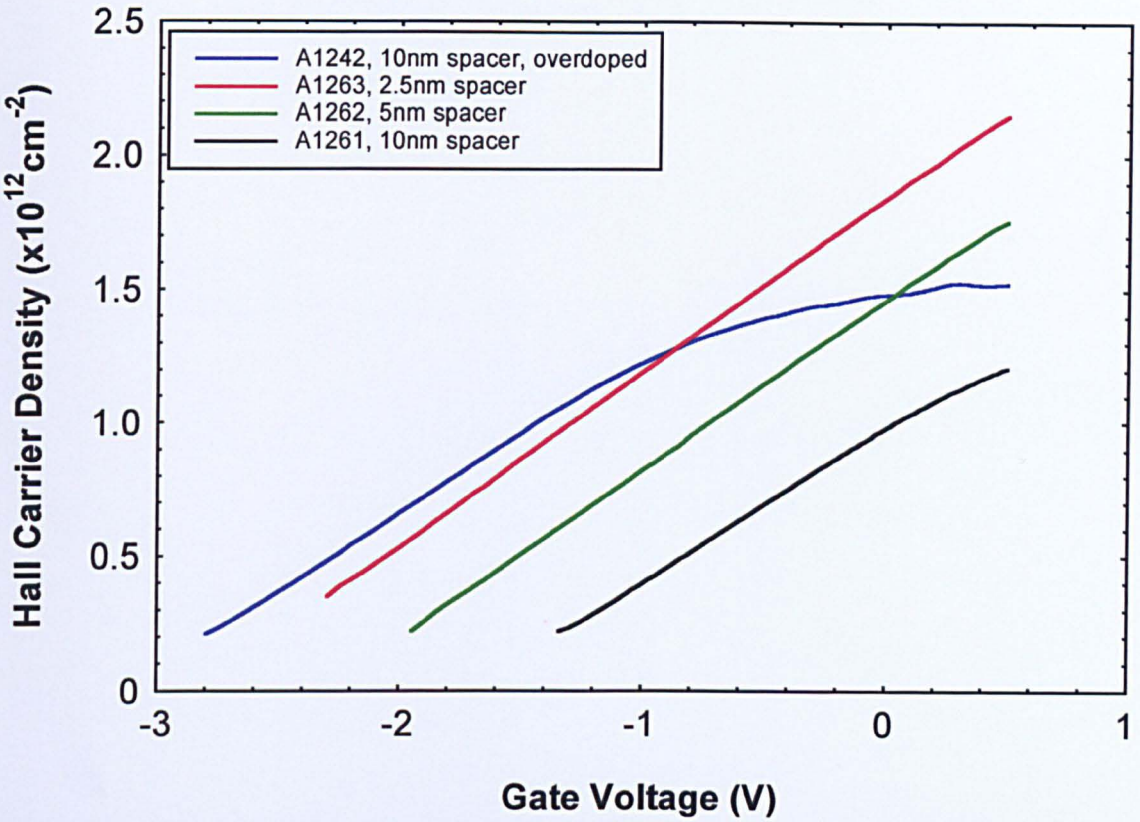


Figure 5.2.2 Plot showing the experimental carrier density depletion at 1.5K from a highly over-doped sample with a 10nm spacer (A1242) along with the depletions of the three sub-critically doped samples A1261, A1262 and A1263. Note the difference in linearity between samples with different levels of over-doping.

As a final point, it is worth emphasising that the critical doping densities are ideal values. In reality, it is likely that these values would have to be exceeded to account for donated electrons getting trapped in various bulk impurity sites, crystal defects, etc. Of course, good quality MBE growth should minimise this.

In summary, the electrostatic model was used successfully to make useful predictions, particularly regarding the electron density in the channel and also the doping density. It would therefore, be an adequate design tool for devices based on these layers.

5.3 Electron Transport at Low Temperatures after a 'Normal' Cool

5.3.1 Introduction

To investigate the elastic scattering factors effectively, the major inelastic scattering component - that due to phonons - must be suppressed. This can be achieved at temperatures close to absolute zero, where the thermal lattice vibrations are minimised. However, phonon scattering should not be ignored for two important reasons. Firstly, these layers will be in devices designed to operate at room temperature, and secondly the elastic strain around the pseudomorphic junctions is bound to affect the distribution of phonons. Device performance could be sensitive to these strain-induced effects.

All the data in this section was acquired 4-terminally after a normal cool at about 1.5K. A normal cool is one in which the gate is shorted to the channel during cooling and the sample may be assumed therefore, to be in electrostatic equilibrium. Low-magnetic field classical Hall measurements were generally used as they were found to be reliable. As many of the elastic scattering mechanisms are electron density dependent, variation of this parameter using a gate potential was necessary. For samples with a shallow channel this potential was varied from about +0.25V to -1 or -2V, whilst samples with a deep channel were subjected to a range +0.5V to -3 or -4V. In either case, the channel was effectively depleted of mobile electrons as signified by a large resistance along the channel.

5.3.2 Observations and Discussion

Fig. 5.3.1 shows the mobility - carrier density data, measured at about 1.5K, for representative samples containing both 20% and 30% indium content in the channel. To avoid confusion in this figure, the next plot, fig. 5.3.2, contains the identical data, but also includes some theoretically derived curves. These are the curves where mobility is only limited by remote ionised impurity (RII) or random alloy scattering. The analytic expressions used are given in equations 2.4.30 and 2.4.48 respectively, in section 2.4. At low carrier density, the RII component is far more likely to be the dominant mobility-limiting factor, while the random alloy component dominates comparatively at high carrier densities. From an examination of fig. 5.3.1, the main observations are as follows:

At low carrier densities (less than about $1x10^{12}cm^{-2}$):

a) All the curves cut the carrier density axis at finite values when these are extrapolated to the zero-mobility cut-off. This implies that a sizeable density of electrons of between 2 and $8x10^{11}cm^{-2}$ is not contributing to mobility in the metallic regime. They are likely to be confined (strongly localised) in the potential energy variations or perturbations of the conduction band-edge. This observation initiated some study of localised transport in these layers and the results are found in section 5.9.

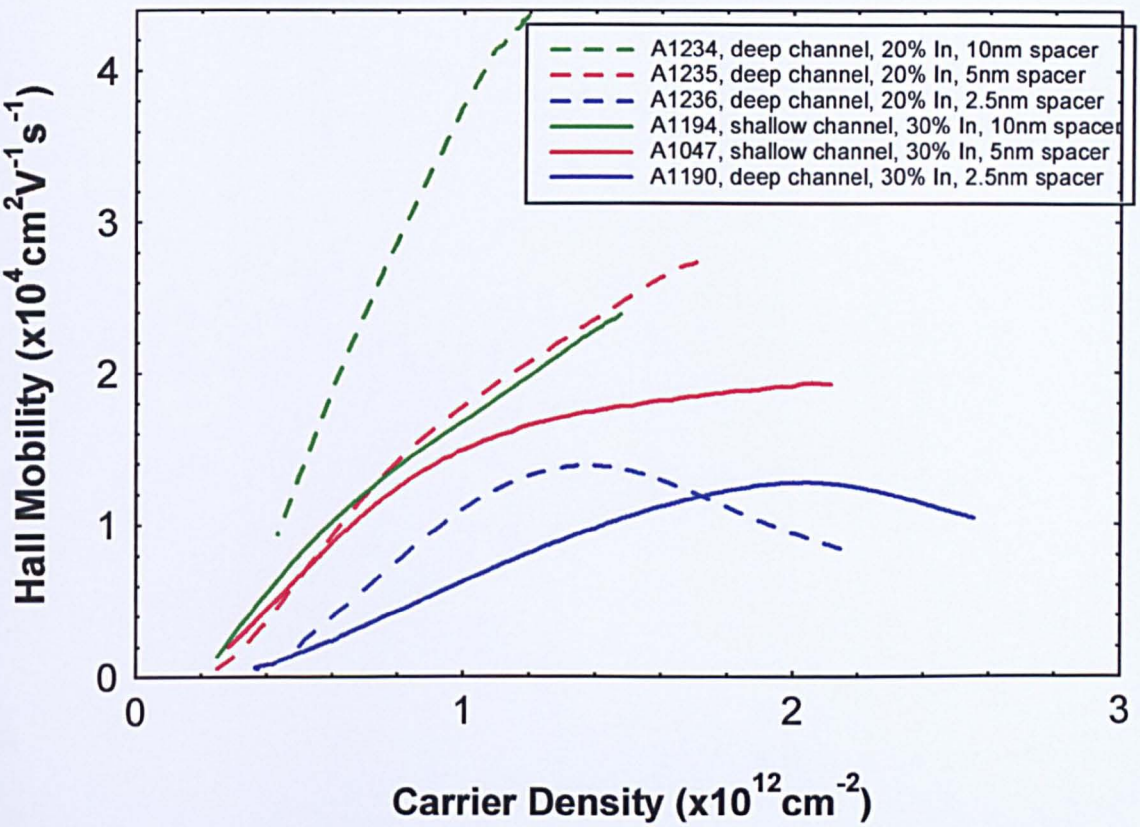


Figure 5.3.1 The Hall mobility for variable carrier density measured at 1.5K after a 'normal' cool. A range of samples with both 20% and 30% indium in the channel was measured; each characterised by three spacer widths (2.5, 5, and 10nm).

Although this will be described in more detail in section 5.9, a simple approximation can be made at this stage. This confining energy can be approximated to the kinetic energy (E_F), associated with the Fermi level, of this low density of electrons. This relation between energy and density is given in equation 5.1.1. Using a value for n of $4\text{x}10^{11}\text{cm}^{-2}$, there is a confining potential energy of about 20meV, see also fig. 2.5.2.

$$E_F \approx \frac{\pi n \hbar^2}{m^*} \qquad (5.1.1)$$

where n is the confined carrier density; and m^* is the effective electron mass.

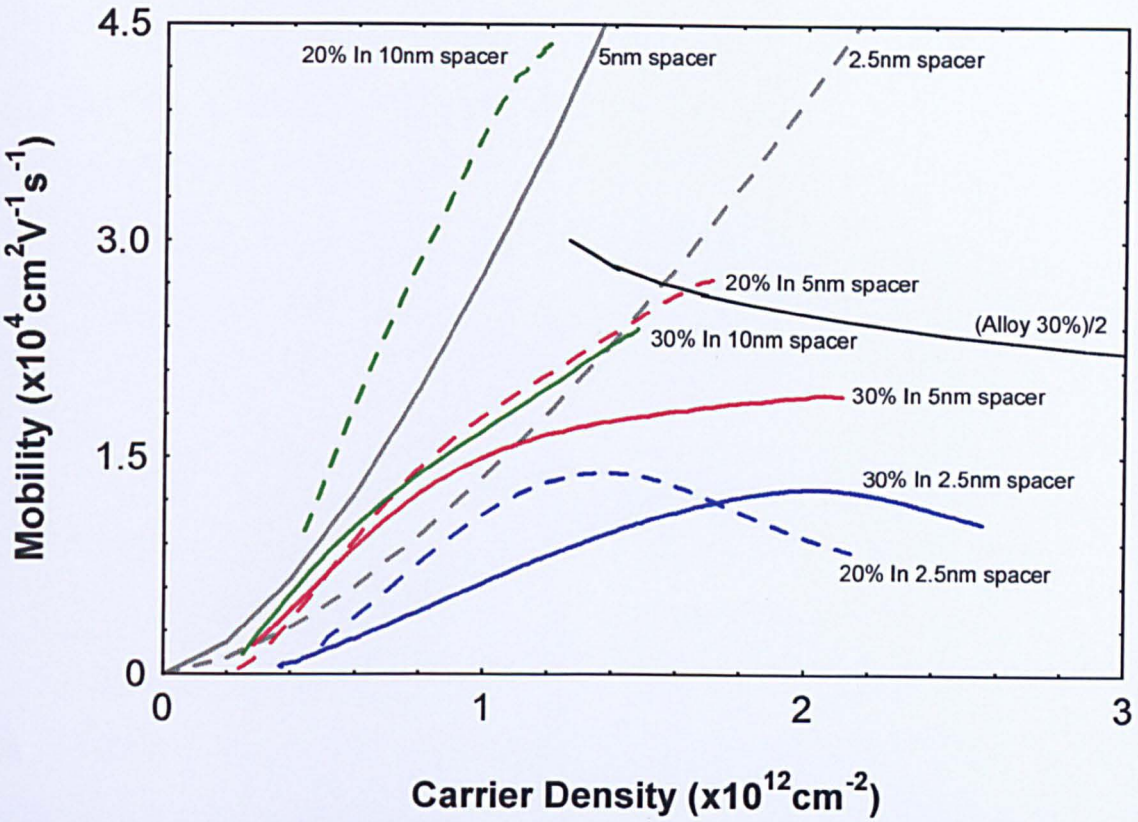


Figure 5.3.2 The same data as presented in fig. 5.3.1 but with the theoretically derived curves for mobility limited by RII and random alloy scattering superimposed. The RII scattering data (gray lines) is spacer dependent and is based on a doping density of $7x10^{12}cm^{-2}$.The data is calculated using uncorrelated donors therefore, it should be regarded as a lower limit (section 2.4). The curve (black line) for 30% random alloy scattering has been divided by 2 to bring it onto the graph. The RII scattering data for a doping density of $3.5x10^{12}cm^{-2}$ is approximately twice the values shown for the higher doping density and hence is even less significant.

b) There exists, for all samples, an approximately linear rise in mobility with carrier density. At low density this indicates that RII scattering may not be the dominant factor. If it were dominant, mobility would satisfy the relationship $\mu \propto n^{3/2}$ as is apparent in the appropriate theoretical curves in fig. 5.3.2. This *3/2*-power trend is often observed when similar measurements are made on GaAs/AlGaAs systems, even with much thicker spacer layers. With such thin spacers, RII scattering would at least be expected to be significant.

c) The spacer width does play some part in the limitation of mobility. It is quite clear from fig. 5.3.1 that, at low carrier density, the sample with the thinnest spacer, for example A1190 (30% indium, 2.5nm spacer), has a significantly lower mobility than the samples with a thicker spacer. This and the observation in *b* suggest that although RII scattering plays an increasingly significant role as the spacer width is reduced, it never dominates scattering for any 30% sample.

d) The samples with a 20% indium content - A1234 (10nm spacer) and A1236 (2.5nm spacer) - have a significantly higher mobility than their 30% counterparts (A1194 and A1190) respectively, at low carrier density. The others with a 5nm spacer have a similar mobility. Based on this observation and the theoretical curves, it can be concluded that the intrinsic indium content of the channel plays a significant and possibly dominant role in the limitation of mobility at low carrier densities.

All the above observations support the conjecture that, at low carrier density, there is substantial scattering originating in the channel. Except at the very lowest carrier densities in samples with the thinnest spacers, this mechanism appears to dominate all the other scattering mechanisms. Indium is known to segregate then aggregate during the growth phase, see sections 1.1.2 and 3.1.3 and references therein. This mechanism will lead to spatial indium concentration fluctuations in the channel. Marsh (1982)[2] demonstrated a model for electron scattering by such proposed clusters, which fitted well with experiment, and gave support to this model of ternary growth in the earlier years. Across the plane of the channel therefore, the energy band-gap of the InGaAs will vary, leading to long-range fluctuations of the conduction band-edge.

The presence of these fluctuations (20meV) was supported earlier by the observation described in *a*. The idea was introduced in section 2.5.3. If the potential fluctuations are large, it should also influence transport at higher density. Another possible source for this scattering behaviour is fluctuations in the physical thickness of the channel layer, which are also likely to be produced by indium segregation. Measurements made when the Fermi level was close to the top of the potential fluctuations, will be discussed in section 5.9.

Evidence from electron microscopy[3] suggests a large dimension of about 100nm to these fluctuations. From an analysis of the extrapolations to zero mobility in fig. 5.3.1, there is little evidence of any significant difference between samples with 20% and 30% channel indium content. Assuming that a channel layer with a 30% indium content will have stronger clustering, and hence stronger perturbations of the conduction band-edge, one might expect that an extrapolation for a 20% sample would cut the carrier density axis at a lower value. In these circumstances, the confining potentials arising from the possible indium concentration fluctuations would be weaker. Of course, determining accurate carrier densities in the strong localisation regime from measurements made in the metallic regime may be inaccurate. More accurate analysis will have to wait for the appropriate discussion in the strong localisation regime. However, a comparison of the mobility of 20% and 30% samples in the low-density metallic regime does suggest that there is a significant difference in the direct influence of the appropriate scattering potentials, whatever their source.

At high carrier densities (greater than about $1x10^{12}cm^{-2}$):

a) At higher carrier densities than signified above, the mobility of the samples with 10nm spacers retains the trend apparent at lower carrier densities. Conversely, all the samples with spacers of 2.5nm and 5nm appear to have mobility limited by an additional factor. The thinnest-spacer samples are undoubtedly most affected therefore this additional factor may have its origin in the doping layer.

b) As the spacer width decreases there is a greater tendency for mobility to peak, with both the thinnest spacer samples A1190 (30% indium) and A1236 (20% indium) showing strong examples (A1263 and A1046 are not shown in the plot, but their data is very similar to that of A1190). Of the two shown, the 20% indium specimen displays a slightly stronger peak, which occurs at a significantly lower carrier density. Later plots, such as fig. 5.5.1, where this sample is studied in isolation, show, in better perspective, the strength of this peak.

Let us assume that it is the electrons penetrating the barrier and interacting with states around the donors that are causing the reduction in carrier mobility. This is in agreement with the observation that the mobility peak is only seen for the thinnest spacers. Figure 5.3.3 shows how the carrier density depletion curves turn over from a linear trend at lower gate voltage (V_g) (more negative) to a flatter trend at the same gate voltage that the mobility peak strengthens. The flattening trend of the $n(V_g)$ curves indicates that carriers are tunnelling into the shallow states of the donor layer and the correlation of the mobility peaking at the same gate voltage suggests strongly that this transfer process is associated with this additional scattering. Intuitively, this behaviour should be higher for the 20% indium sample, when compared with the

equivalent 30% sample, because its conduction band-edge energy offset at the channel-spacer boundary is of smaller magnitude. Its energy barrier is only 360meV compared with about 430meV of the 30% sample. It is also more strongly overdoped relative to the critical doping density (N_δ) introduced in section 2.1 and discussed in section 5.2.

To be more certain of the actual conduction band-edge energy profile and the wavefunction distribution, a self-consistent analysis, like the one mentioned in section 1.2.3, was carried out. This showed wavefunction penetration of the doping layer for the thinnest spacer samples, and that the energy gradients of the conduction band-edge over the spacer and channel layer are somewhat less for a 20% indium sample when compared with a 30% specimen. Therefore, the electric fields either side of the spacer layer will be smaller in magnitude. Therefore, the scattering mechanism causing peaking should come in at a lower carrier density than that for the equivalent 30% specimen. Fig. 5.3.1 shows that the mobility peak for a 20% sample (A1236) occurs at a carrier density of only $1.4\text{x}10^{12}\text{cm}^{-2}$, compared with about $2\text{x}10^{12}\text{cm}^{-2}$ for the 30% sample (A1190). As stated above, the degree of over-doping, if any, can play a significant part. The 20% indium sample was more overdoped than the 30% sample, and this may explain the slightly greater strength of its peak. The additional scattering is so acute in the 20% sample that, at a carrier density above $1.7\text{x}10^{12}\text{cm}^{-2}$, its mobility is now exceeded by that of the 30% sample despite the stronger intrinsic channel scattering of the latter discussed earlier.

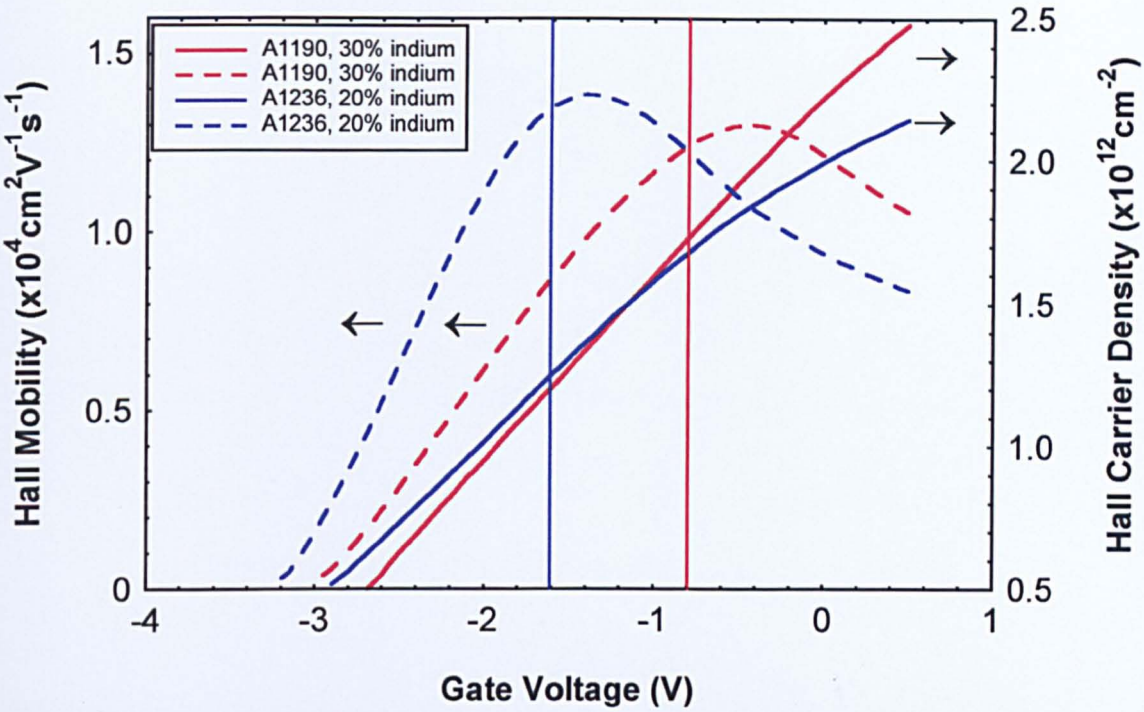


Figure 5.3.3 The relationship between the Hall carrier density depletion (solid line) and mobility (dashed line) with gate voltage at 1.5K for samples A1236 (20% indium, 2.5nm spacer) and A1190 (30% indium, 2.5nm spacer). Note the common voltage in each case (A1236 -1.6V) (A1190 -0.8V) at which the mobility begins to strongly peak and the depletion begins to flatten.

The longitudinal resistance in the 2.5nm spacer samples was observed to pass through a minimum as the channel was being depleted. It can be shown mathematically[‡] that if the resistance has this behaviour, then parallel conduction can be eliminated as a possible mechanism for the peaking. This was also confirmed experimentally with the absence of a background B^2 behaviour to the appropriate SdH plots at low temperature. The significance of random alloy scattering is discussed in *c* below. Therefore, some influence in the doping layer or thereabouts is likely to be responsible for this peaking characteristic. This would be acting directly on the electrons in the tail of the wavefunction. Of course, these electrons must exist for a sufficient time in the doping region for the effect to act. One process being considered at the time of

[‡] A.R. Long - Private Communication.

writing[‡] is that the channel electrons, which penetrate a thin spacer, can interact with the states there, before making a transition back to the channel - a form of resonant scattering. This type of scattering would be expected to degrade the overall mobility quite significantly. Much more will be discussed relating to the peaks in later sections such as when mobility anisotropy and a comparison of quantum and transport mobility are considered. Other interesting phenomena related to the mobility peak also appear when the temperature is varied, see section 5.5.

c) Random alloy scattering is characterised in theory by a shallow decline in mobility with increasing carrier density, see fig. 5.3.2. The sharpness of the mobility reduction past the peak, particularly for the 20% indium sample, is strong evidence that **random** alloy scattering is not significant in forming this peak. The theoretical line representing the 20% indium content is off the top of the graph. However, it should be remarked that the magnitude of the alloy scattering potential is difficult to quantify, therefore a degree of uncertainty in its exact position is inevitable. It is particularly important when fitting theory for random alloy scattering to experimental data that clustering is not ignored as errors are bound to arise.

Finally, in studies of the inversion layers in Si-MOSFETs made in the 1970s[(4)], mobility peaks arising then were thought due to interface roughness scattering at the semiconductor-oxide interface. This factor was said to dominate mobility at densities beyond the peak, superceding RII, or its equivalent scattering, which dominates before the peak. This mechanism cannot be true in this case as the mobility peak only occurs for the thinnest spacers and moreover, as stated in *b* above, it is

[‡] J.H. Davies – Private Communication

correlated with the gate voltage at which charge enters the doping layer. In addition, the interface between the inversion layer and the spacer is grown before the remainder of the spacer width and the subsequent incorporation of the donors. This would mean that any effect due to the interface, for example roughness, must be independent of these.

5.3.3 Section Summary

At low carrier densities in the metallic regime, the mobility is dominated in all samples by an intrinsic scattering factor arising in the channel. RII scattering further limits mobility in samples with thinner spacers (2.5nm). The intrinsic factor is possibly ascribed to the long-range potentials arising from the indium concentration fluctuations or channel thickness fluctuations. Spatial potential perturbations of the conduction band-edge of the channel and hence band-gap will exist, caused by the spatial clustering of indium in the channel layer. In the metallic regime, the channel Fermi level would lie above such a potential 'landscape', but the wavefunction of the channel electrons would, to a greater or lesser extent, be influenced and hence scattered by their presence. As the Fermi level falls with decreasing carrier density, it would eventually sink into the potential fluctuations. Here, significant entrapment of the electrons would be expected at low temperatures, and this is confirmed by the large carrier densities still present when the metallic mobility is extrapolated to zero.

At high carrier densities the influence of the cluster potentials should lessen but not necessarily become insignificant. For the thickest-spacer samples, scattering from this source continues to be dominant, but for thinner-spacer samples another factor emerges. There is a strong

additional spacer-dependent factor, which limits mobility at these densities with a characteristic peak emerging for the thinnest-spacer samples. Direct scattering of the tail of the wavefunction of the channel electrons by factors within the donor layer is thought to be the cause of this behaviour. Evidence of RII scattering exists at all carrier densities if the spacer width allows, but is never thought to dominate the scattering characteristic of any sample at any density. Random alloy scattering, even at high densities and allowing for the uncertainty in its scattering potential, is thought to be insignificant in these samples. To make further advances, additional evidence will be required and this will be discussed in subsequent sections beginning with the evidence revealed by bias cooling.

5.4 Electron Transport at Low Temperatures after a Bias Cool

5.4.1 Introduction

To investigate more fully the effect of RII scattering and its variation with spacer thickness, a selection of the earlier samples was subjected to bias cooling. As stated in sections 1.4 and 2.3, this process can alter the electron concentrations trapped in the deep-donor or *DX* centres. The occupation of the donor sites is correlated, which will reduce the effect of the scattering field of the ionised donors. Bias cooling is simply one of the methods which can vary the magnitude of this correlation.

The percentage occupation of these *DX* centres can be adjusted between about 20% of the total donor density and essentially zero by applying a variable gate potential during the sample-cooling stage. This potential must be applied above the *DX* 'freezing' temperature thus allowing thermal energy to activate the electrons out of the centres. The active electrons can be attracted to or repelled from the donor layer prior to cooling, depending on the sign of the potential. In this way, the final occupation of the deep centres after cooling under bias can be controlled to some degree and hence permit 'tuning' of the correlation arising in that layer.

Once the bias is released at low temperature, electrons can only occupy the channel or the shallow states of the donor layer from where they can easily be removed during measurement. Shubnikov-de Haas plots confirmed that in all samples measured, neither a second subband

occupation nor a parallel conductance 'pathway' was detected under any low-temperature experimental conditions. This was despite the increase in the electron occupation of both the states of the channel and the shallow states of the donor layer after cooling with negative gate bias. Therefore, most of the mobility measurements were acquired using the low-magnetic field Hall effect. SdH measurements were only used to obtain some quantum mobility values where appropriate, which could reveal the changing emphasis of RII scattering as the correlation changed.

5.4.2 Observations and Discussion

The carrier density dependent Hall mobilities for a selection of samples is shown in figs.5.4.1 to 5.4.3. They were acquired at the base temperature of 1.5K, subsequent to cooling under a range of gate potentials applied from between about 240K and 1.5K. The most notable features displayed in these and other data are as follows:

a) At a given carrier density, the mobility falls with an increasingly negative cooling bias. This is entirely as expected. A positive gate bias applied during the cool would increase the probability of more electrons being found in the doping layer region. The attractive or confining gate potential would discourage electrons sited there re-locating to the other potential minimum - the channel. The increased presence of the electrons in the doping layer would enhance their likelihood of entrapment in the local *DX* centres as the sample cools below their freezing temperature. Below this temperature - about 70K to

150K for $Al_{0.3}Ga_{0.7}As$ - electrostatic energy, derived from the gate, is insufficient to remove the electrons, although it can be achieved with appropriate illumination. This increased occupation of the *DX* centres would enhance correlation of the donor layer charge, and smooth out the resulting potential perturbations of the conduction band-edge in the channel. Another equivalent view would invoke the formation of dipoles forming between the local positive and the additional negative charge, and this would somewhat reduce the scattering field of the ionised

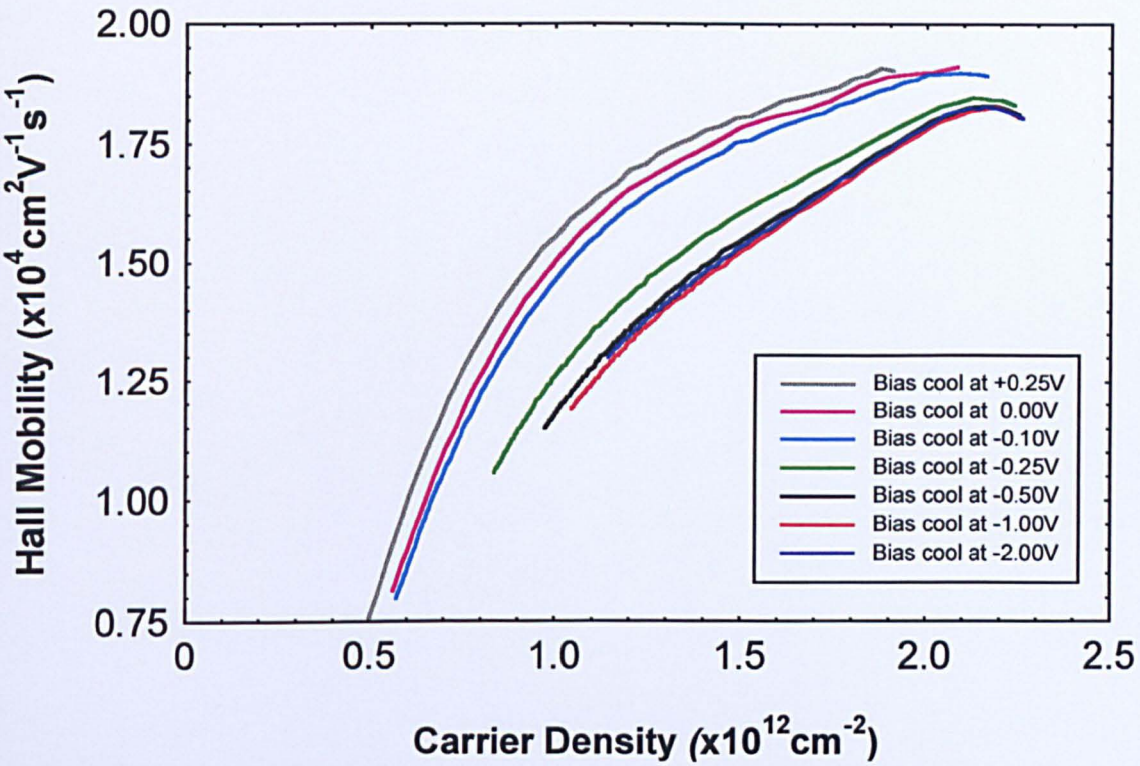


Figure 5.4.1 The variation of Hall mobility with carrier density for sample A1047 (30% indium, 5nm spacer) at 1.5K, when cooled under a variable bias. Note the reduction in mobility with an increasingly negative bias cool.

donors with respect to isolated charges, see also section 2.3. The above ideas would also suggest that there would be a measured reduction in the carrier density of the channel at zero gate voltage associated with the

increased mobility at a given carrier density. This is observed in all the sample plots.

The application of increasingly negative gate bias during cooling increases the probability of the electrons in the doping layer, and indeed the channel, being repelled towards the substrate. Therefore, fewer

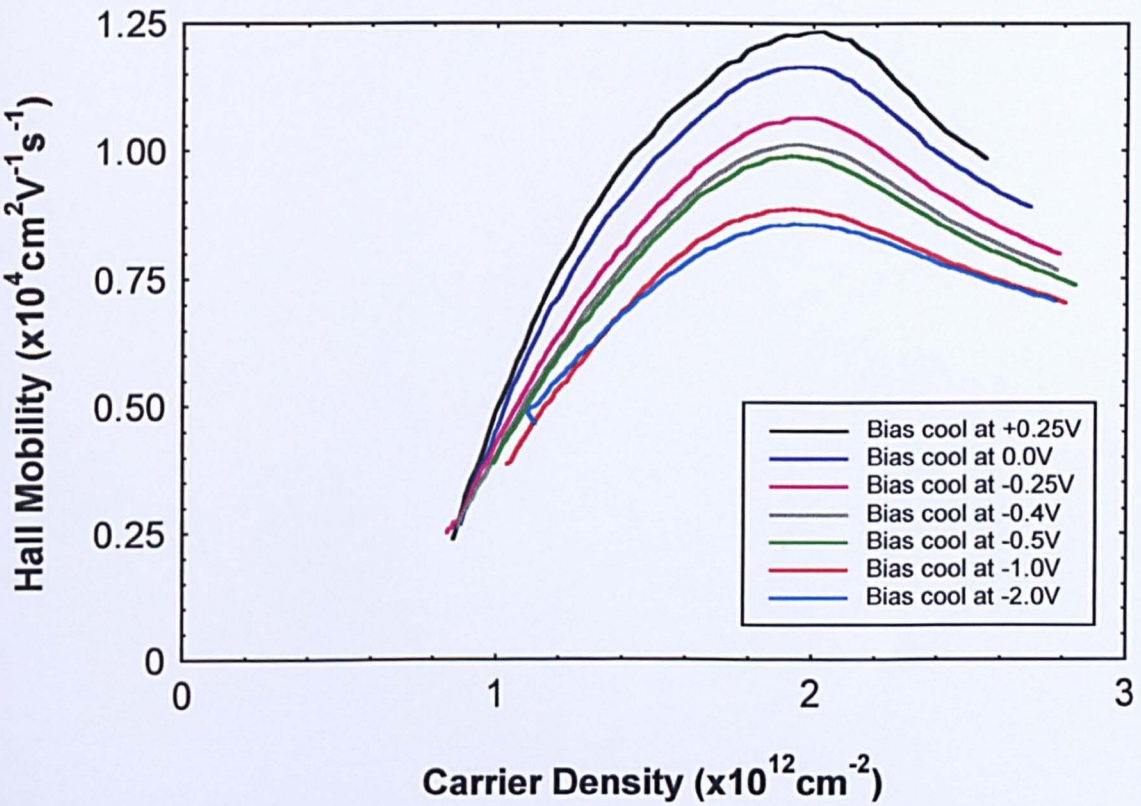


Figure 5.4.2 The variation of Hall mobility with carrier density for sample A1046 (30% indium, 2.5nm spacer) at 1.5K, when cooled under a variable bias. Note the reduction in mobility with an increasingly negative bias cool. The peaking characteristic gets weaker as well.

electrons would condense into the *DX* centres as the temperature passes through the critical value. At base temperature, when the cooling bias is released, the electrons can return to the potential minima of the layers. Although some of the electrons return to the doping layer, they can no

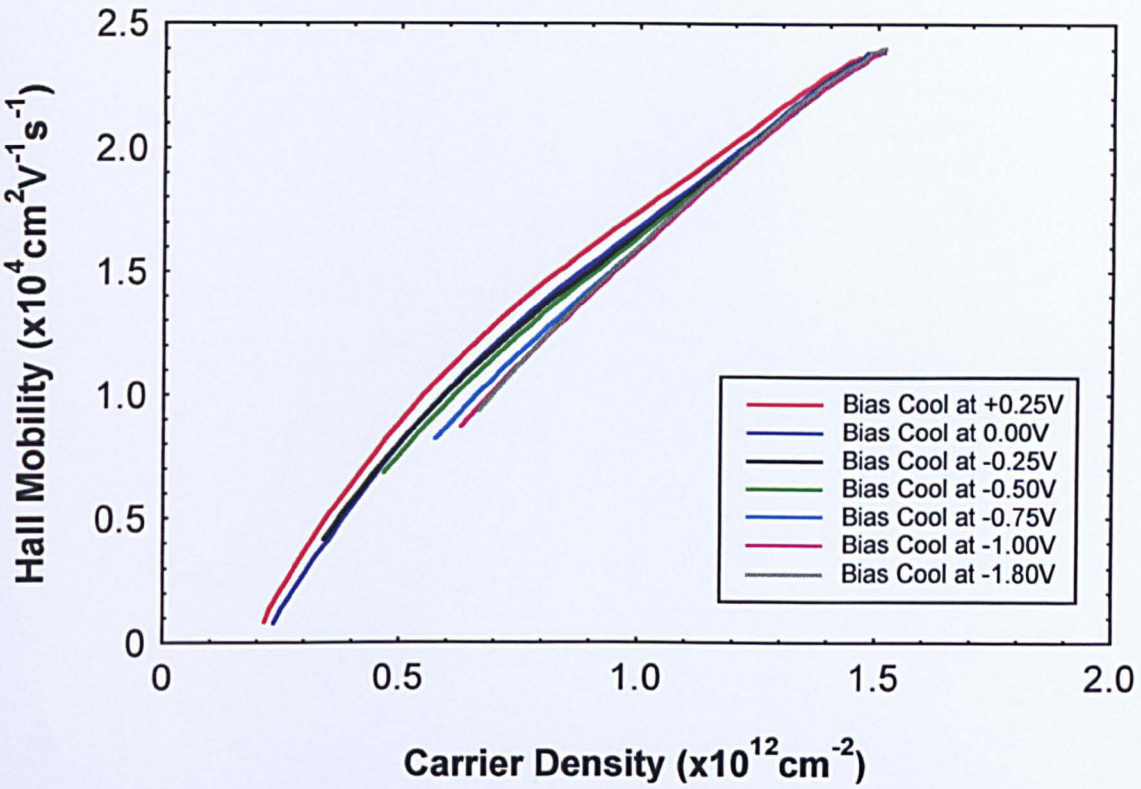


Figure 5.4.3 The variation of Hall mobility with carrier density for sample A1194 (30% indium, 10nm spacer) at 1.5K, when cooled under a variable bias. Note there is little, if any, reduction in mobility with an increasingly negative bias cool, except at lower carrier density.

longer occupy the *DX* levels. They have insufficient thermal energy to enter them and hence, must remain with in local shallow states, see section 1.4. From there, they are easily removed by the gate bias subsequently applied in measurement and when this happens any smoothing of the RII potential is reduced correspondingly. The resultant undulating potential topography at the channel will increase RII scattering and reduce mobility. This is also observed in all plots, although there is a tendency for the effect to saturate at the more negative bias cools. This will be shown in greater detail later in fig. 5.4.4.

b) At higher carrier density, all samples display some change in the mobility with changing donor layer correlation, indicating that RII scattering, whilst not dominant, is not negligible either in most of these layers. Not surprisingly, the sample (A1194) with the thickest spacer (fig. 5.4.3) was affected least, particularly at higher density. In this sample, the channel electrons are much more remote from the donor layer, and hence the influence of any change in its charge state after a bias cool will be relatively much less. At lower carrier density, a greater reduction is apparent due to the removal of the electrons occupying the shallow states of the donor layer. The screening of the ionised donors by these electrons will therefore be much reduced. This particular sample was also over-doped therefore, even more electrons would have to be removed from these shallow states. Less obvious was the relative similarity of the magnitude of the mobility changes in the samples (A1046 and A1047) with 2.5 and 5nm spacer widths respectively (figs. 5.4.1 and 5.4.2). However, if the thinnest spacer sample is already severely mobility-limited by the direct-donor scattering mechanism referred to previously, then the bigger changes expected will be somewhat masked. The electrons in the channel, particularly at higher densities, would also be expected to provide additional screening from the remote scattering effect.

c) There is, in general, a non-linear change in mobility with variable bias cool voltage at a given carrier density. Fig. 5.4.4 shows this relation for several carrier densities in the sample A1047, which contains a shallow channel with 30% indium content and a 5nm spacer.

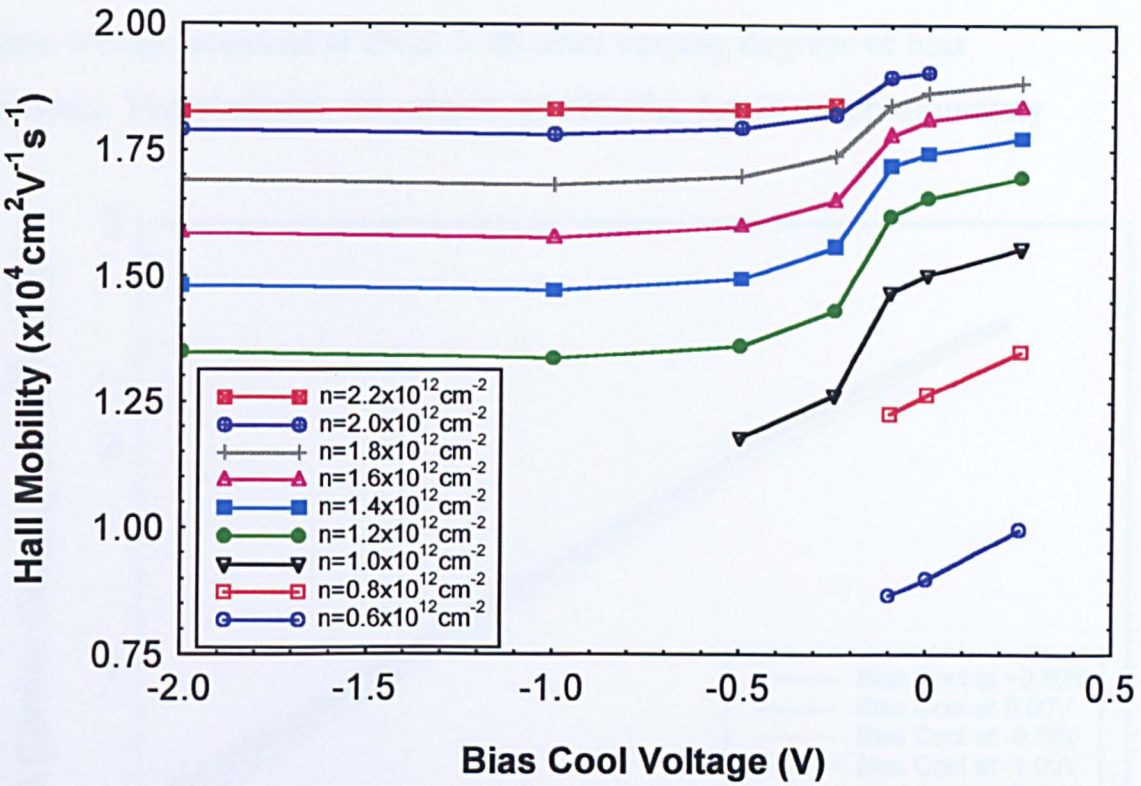


Figure 5.4.4 The variation of Hall mobility with the voltage applied during cooling for sample A1047. The data was acquired at 1.5K at several denoted carrier densities. Note that most of the change occurs over a very narrow range in the bias cool voltage, with a saturation at the more negative bias cools.

In this sample and others, a clear step in mobility evolves with increasingly negative bias cooling. In other words, most of the change in the mobility occurs over a very narrow range around a bias cool voltage of about -0.20V, with little or no change over the remaining range. This implies that there is a relatively sudden switch of electron occupation of the deep donor or *DX* centres in this locality. The higher mobility 'level' is characterised by deep donor occupation and strong correlation, whilst the lower is distinguished by effectively low *DX* level occupation and a random RII potential.

d) In addition to the mobility, the trends in carrier density can provide useful information. Figs. 5.4.5 and 5.4.6 show, for selected

samples (A1190 and A1194), the depletion of the carrier density with gate voltage acquired at about 1.5K after varying degrees of bias cooling. The depletion for sample A1190 (fig. 5.4.5) is approximately

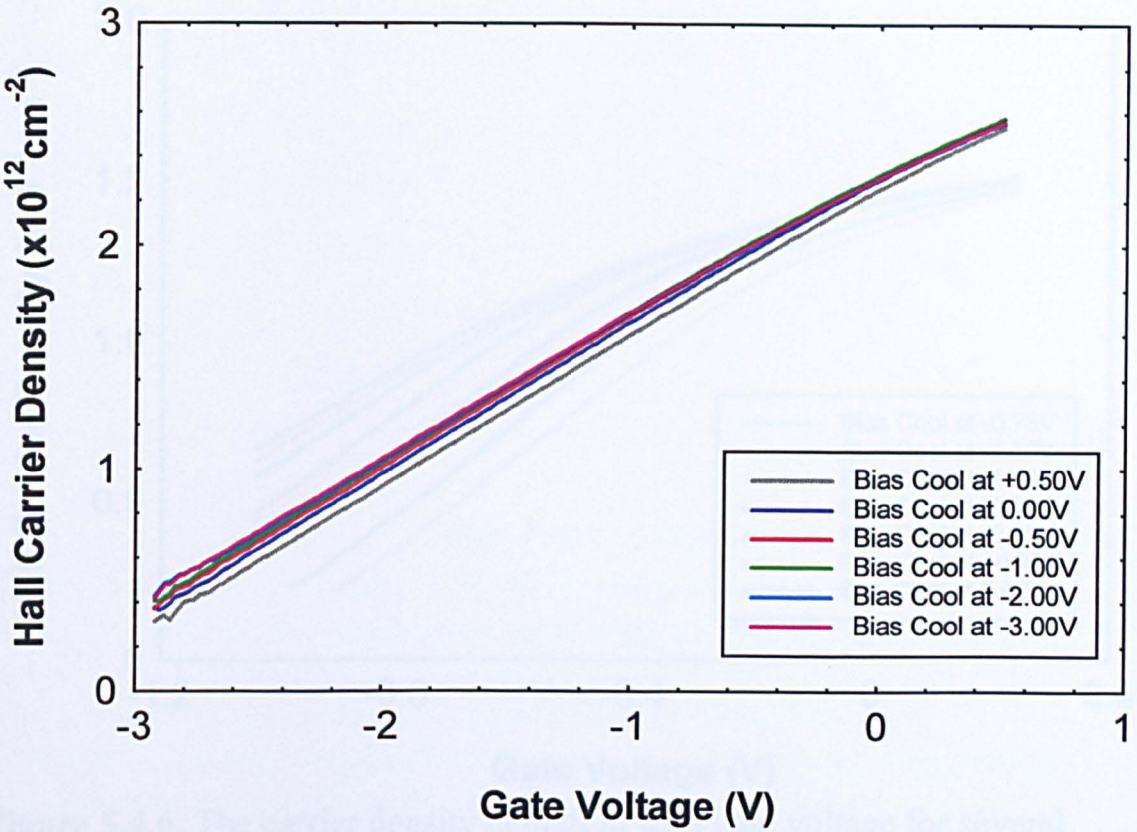


Figure 5.4.5 The carrier density depletion with gate voltage for sample A1190 over several bias cool voltages. The data was acquired at 1.5K.

linear over the whole gate voltage range, and does not vary much with bias cooling. On the other hand, in sample A1194 (fig. 5.4.6) there is distinct curvature in the depletion around zero gate bias and significant variation with bias cooling. As stated earlier, the presence of a flatter depletion curve around zero gate voltage tends to implicate the over-population of the shallow states of the donor layer arising from over-doping. After negative bias cooling there are even more electrons in shallow states at zero gate voltage. These need to be depleted before the channel density drops markedly. From the results of the electrostatic model in section 5.2, the 10nm spacer samples, such as A1194, were

invariably more over-doped than the thinner spacer samples such as A1190. Sample A1194 in fig. 5.4.6 illustrates this effect strongly, while

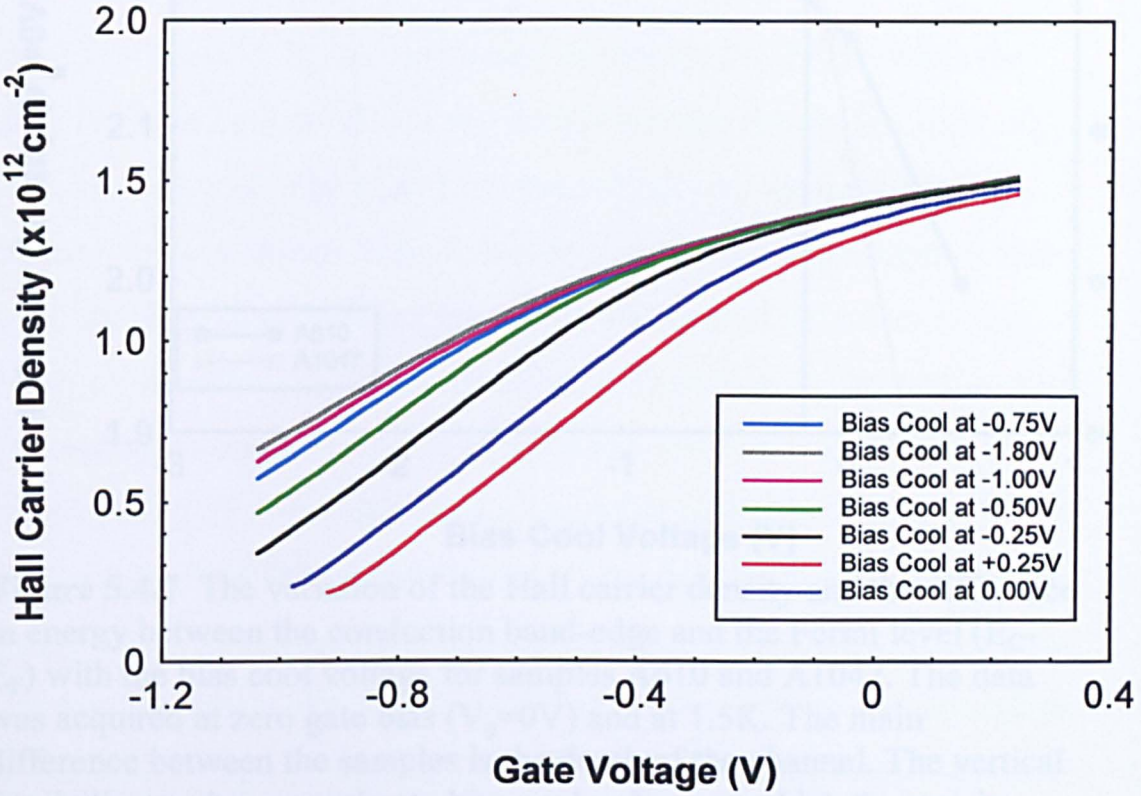


Figure 5.4.6 The carrier density depletion with gate voltage for several bias cool voltages for sample A1194. The data was acquired at 1.5K.

the change is small for sample A1190, in fig. 5.4.5. Indeed, for the former sample linear depletion barely evolves at all at the most negative bias cool. The intermediate 5nm spacer samples (not shown) display an intermediate behaviour. Negative bias cooling also increases the channel density hence greater negative gate threshold voltages would be expected, and are observed.

e) If the low-temperature carrier density at zero gate voltage is plotted against the bias cool gate voltage, some interesting features emerge. Such a plot is shown in fig. 5.4.7 for two samples, A1047 and

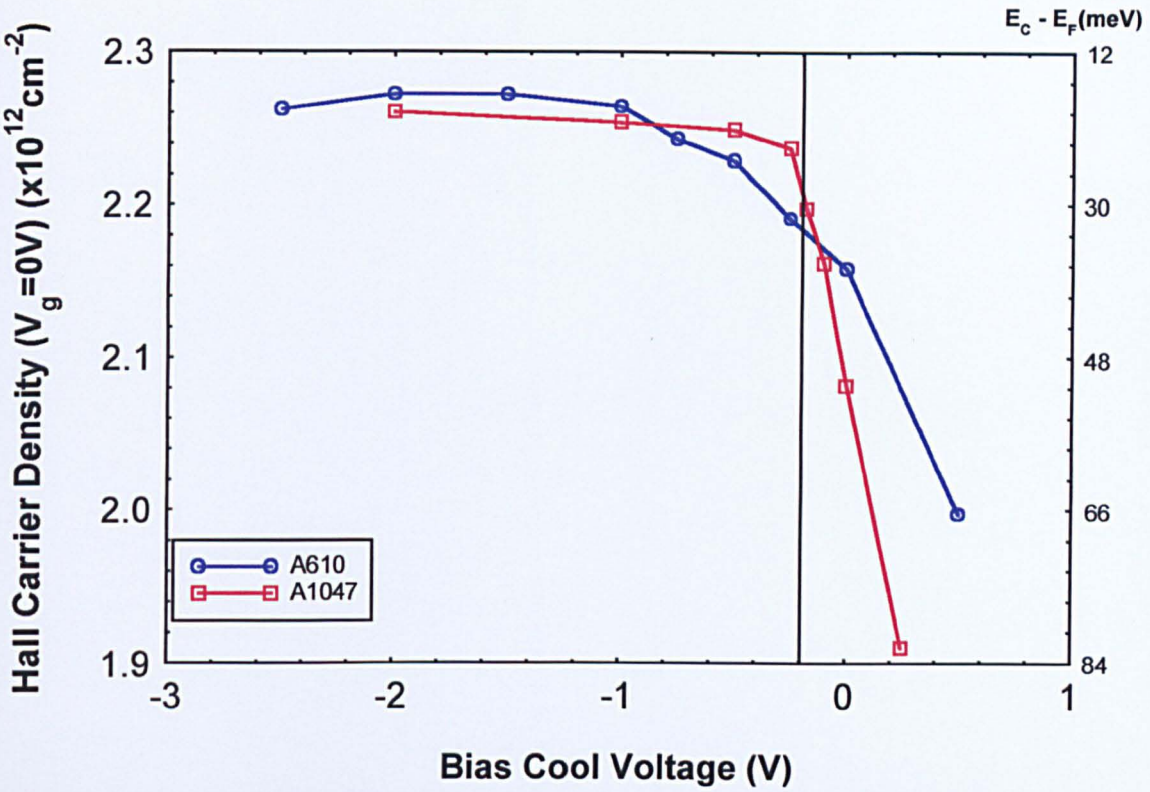


Figure 5.4.7 The variation of the Hall carrier density and the difference in energy between the conduction band-edge and the Fermi level (E_C-E_F) with the bias cool voltage for samples A610 and A1047. The data was acquired at zero gate bias (V_g=0V) and at 1.5K. The main difference between the samples is the depth of the channel. The vertical line indicates the approximate bias cool voltage at which the step in mobility occurred in fig. 5.4.4.

A610. These samples only vary in the depth of their channel - A1047 being the shallower. Both samples show an increase in channel density followed by saturation, but the trend for A1047 is more discontinuous with a prominent 'elbow'. Note that for A1047, the saturation in carrier density, below a bias cool voltage of about -0.20V, corresponds to the saturation in mobility in fig. 5.4.4 for the same specimen. Also, the bias cool voltage of the 'elbow' in fig. 5.4.7 corresponds to that of the mobility step in fig. 5.4.4.

f) More intriguingly, however, if the deep donor energy from equation 2.1.9 is re-expressed in terms of the carrier density, that is:

$$E_{dd} = \Delta E_C - \frac{ne^2}{\varepsilon_0}\left[\frac{s}{\varepsilon_{spacer}} + \frac{a}{\varepsilon_{channel}}\right] \qquad (5.4.1)$$

then the values for E_{dd} (or the energy difference (E_C-E_F) between the effective Fermi energy (E_F) at the donors and the conduction band edge (E_C), see below) can be placed correspondingly on the right-hand ordinate axis, as shown. Note however, that under a large negative bias cool, the Fermi level will not pin on the unoccupied deep donors, but rather on the energy of the shallow states. Therefore, as the cooling bias becomes more negative, the pinning energy relative to the conduction band-edge (E_C-E_F) decreases as the transition takes place from occupation of the deep donors to that of the shallow states. This pinned energy value changes from about 70meV (the value largely accepted for electrons trapped in the deep donors here) to about 20meV at saturation, which is an accepted value for the shallow donors in this material. This substantiates the claim that the change in donor layer correlation, and hence that of channel mobility, is made possible by the removal of electrons from the deep donor states to the shallow states from where they can be easily removed electrostatically.

g) In section 5.3, it was established that the mobility over much of the carrier concentration range is lower than expected for RII scattering. If an effective mobility ($\mu_{channel}$) is assigned to the scattering in the channel alone, then using Matthiessen's rule (equation 5.4.2), this mobility component can be combined with one for the RII scattering by adding the reciprocal values, proportional to the scattering rates, to obtain the overall reciprocal mobility ($1/\mu$):

$$\frac{1}{\mu} = \frac{1}{\mu_{channel}} + \frac{1}{\mu_{RII}} \tag{5.4.2}$$

The RII mobility component can be correlated ($\mu_{correlated}$) or uncorrelated ($\mu_{uncorrelated}$) depending on the bias cool voltage and hence, the overall mobility in these conditions ($\mu_{Negative\ Bias\ Cool}$ or $\mu_{Positive\ Bias\ Cool}$) will be affected. This is shown in the following two equations:

$$\frac{1}{\mu_{NegativeBiasCool}} = \frac{1}{\mu_{channel}} + \frac{1}{\mu_{uncorrelated}} \tag{5.4.3}$$

$$\frac{1}{\mu_{PositiveBiasCool}} = \frac{1}{\mu_{channel}} + \frac{1}{\mu_{correlated}} \tag{5.4.4}$$

If equation 5.4.4 is subtracted from equation 5.4.3, the channel mobility term can be eliminated. This leaves a 'residual mobility' ($\mu_{residual}$):

$$\frac{1}{\mu_{residual}} = \frac{1}{\mu_{uncorrelated}} - \frac{1}{\mu_{correlated}} \tag{5.4.5}$$

This residual mobility is plotted for samples A1047 and A1046 in fig. 5.4.8 along with the corresponding relations for the uncorrelated RII scattering component derived theoretically. The only effective difference between the experimentally and theoretically derived data now is that the former still contains the mobility enhancing correlation established in the doping layer. (The experimental data was acquired after a normal or zero-bias cool.).

If $\mu_{correlated}$ is much greater than $\mu_{uncorrelated}$, then $\mu_{residual}$ is approximately equal to $\mu_{uncorrelated}$, and $\mu_{residual}$ (dashed curves) should coincide with the uncorrelated RII scattering theory (solid curves) in fig. 5.4.8. If $\mu_{correlated}$ is greater than or approximately equal to $\mu_{uncorrelated}$, then the theory will form a lower bound. This is what is observed for the 5nm spacer sample, implying that the RII scattering mechanism operates

as expected here. However, for the 2.5nm spacer sample, the residual mobility line is already below that of the theory for RII scattering, therefore there must be an additional factor, and suggests that negative bias cooling has an another effect, possibly opening up further resonant scattering 'channels' – a mechanism suggested in section 5.3.

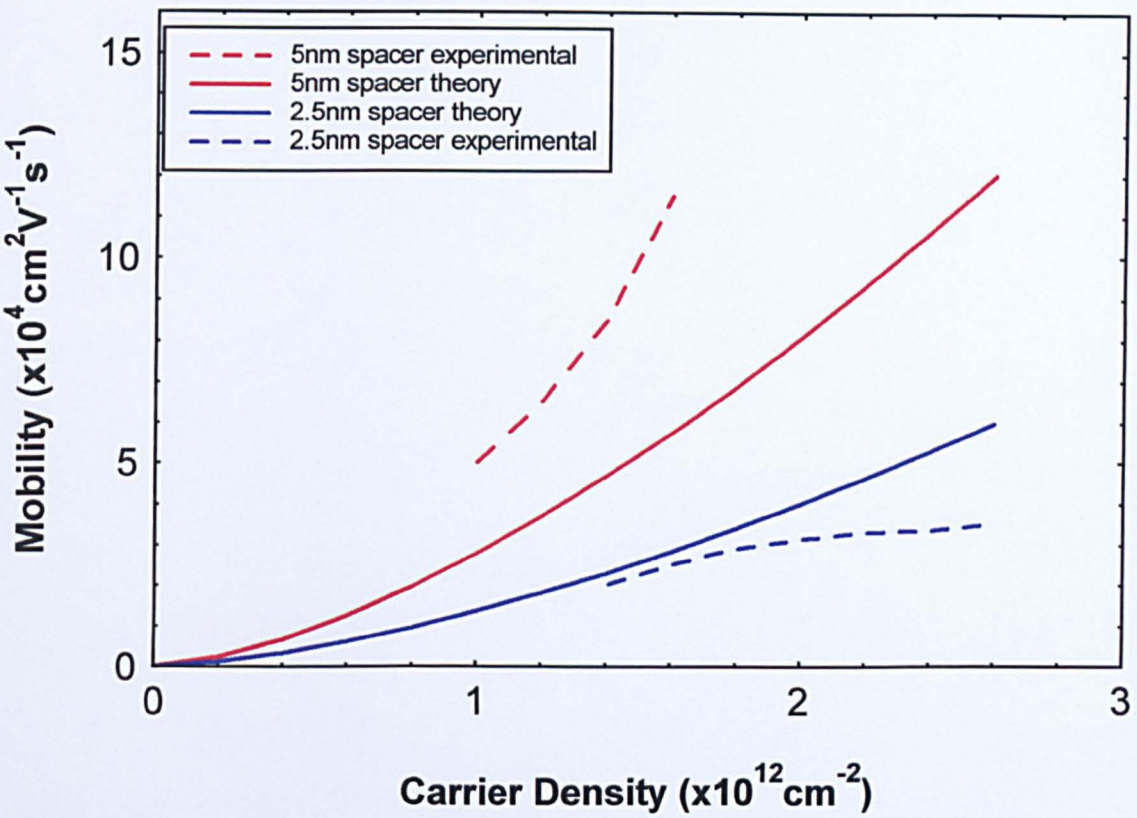


Figure 5.4.8 A plot of the experimentally derived 'residual' mobility against carrier density for samples A1046 (30% indium, 2.5nm spacer) and A1047 (30% indium, 5nm spacer), where the channel component has been 'eliminated' using Matthiessen's rule. Also shown are the corresponding theoretically derived mobility curves for uncorrelated RII scattering. See text for details.

Calculation of quantum mobility was also carried out on bias cooled samples but these results will be discussed in the appropriate section on quantum mobility (section 5.8).

5.4.3 Sectional Summary

The correlations established between the electrons trapped in the deep donor centres of the doping layer were 'tuned' by the process of cooling under a gate bias. This was done to determine the relative effect of the RII scattering component in layers of this type. The application of a positive bias during cooling enhanced the occupation of such centres, while it was discouraged under a negative potential. Therefore, variation of this potential varied the occupation of the donor centres and the extent of the donor-layer correlation. All the samples tested with this method displayed changes in the mobility (transport mobility) of the channel electrons arising from this correlation tuning. There was a tendency for the mobility to change in an approximate step-like manner over a narrow range of bias cool voltage with a saturated low-mobility level at higher negative values. At this saturation, the channel electrons were subjected to the full scattering effect of the uncorrelated ionised donors. A similar saturation is found in the carrier density of the channel over the same range, with the corresponding occupational state energy of the donor layer passing through that for deep donors and saturating at a value commensurate with the shallow states.

The sample specific effect of RII scattering can be quantified by using Matthiessen's Rule and the magnitude of the correlation (mobility changes) exposed in the bias cool data. In this way, the scattering of channel origin and the donor layer correlation can be 'eliminated' to permit a valid comparison with the theoretically derived values for uncorrelated RII scattering. It was found that in those samples tested, what factor remained in the samples of intermediate spacer (5nm) was adequately accounted for by RII scattering. However, an additional

factor had to be introduced (direct donor scattering) to account for the data for the samples with the thinnest spacer (2.5nm). Particularly at high carrier density, RII scattering had little effect on the specimens with the thickest spacers (10nm).

5.5 Electron Transport as a Function of Temperature

5.5.1 Introduction

In the original strategy for this project, variable temperature work in the metallic regime was not specified as it was thought that there would be little to gain. Unlike GaAs/AlGaAs systems, any temperature variation of the relevant parameters, such as mobility, was thought likely to be small - certainly up to about 100K.[5] Scattering by intrinsic channel components, which is fairly insensitive to temperature, was anticipated to dominate any phonon scattering arising from the increasing thermal energy in the temperature range considered. Whilst this is essentially true, an unrelated development altered this strategy. As one particular sample (A1236) was being normally cooled to the base temperature of 1.5K, it was observed that the longitudinal resistance (R_{xx}) was strongly non-monotonic. This resistance passed through a minimum at about 40K, then increased substantially (by 10%) while the temperature dropped further to base temperature.

Prompted by this observation, and the desire to account for the bahaviour of the conductivity as the strongly localised regime is approached, the following measurements were initiated. For each of a range of temperatures, data was acquired 4-terminally using, as before, the low-magnetic field classical Hall technique. This was considered acceptable at lower temperatures where no second subband occupancy or parallel conduction confused the issue. However at higher temperatures, the real possibility of parallel conduction should be

considered. Some SdH data was also acquired mainly at high depletion in the metallic regime, but also occasionally at full channel density to complement the Hall data. A temperature range of between 1.5K and about 130K was chosen, the upper limit being set to avoid the complications arising due to the thermal activity of the *DX* centres around and above this value. Cooling was undertaken normally, although had more time been available, a bias could also have been applied to hopefully reveal more detail. The results reported in this section will be restricted to those acquired using the cryostat, with those taken at room temperature discussed in section 5.7.

5.5.2 Observations and Discussion

The relation between Hall mobility and carrier density for five samples (A1236, A1190, A1235, A1234 and A1242) was determined over a range of temperatures. Of these sets of data, three are shown in figs. 5.5.1 to 5.5.4 (two relate to the same sample). The data for the sample A1236 with a 20% channel indium content and a 2.5nm spacer (fig. 5.5.1) is considered first. Many of the features of this sample are duplicated in the other samples, therefore many of the observations below also apply to them. However, this sample contains additional features. As discussed in the introduction, a distinct reversal in the trend of the resistance occurs at about 40K, therefore we consider firstly the mobility above that temperature. The observations include:

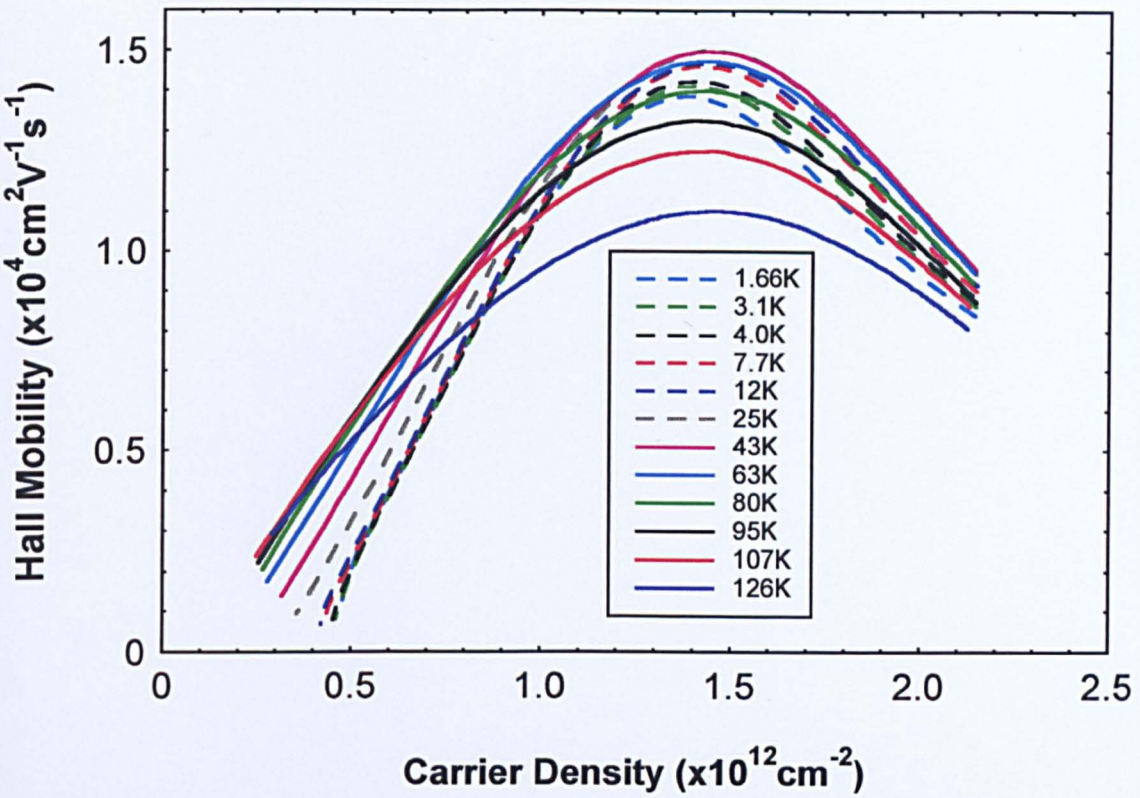


Figure 5.5.1 The variation of Hall mobility with carrier density at variable temperature after a normal cool for sample A1236 (20% indium, 2.5nm spacer). The top temperature was restricted to avoid complications with *DX* centre occupancy.

a) At the highest temperatures considered (>100K), scattering due to phonons becomes the dominant factor. The mobility at room temperature will be shown in section 5.7, but for comparison its maximum value is just less than $5x10^3cm^2V^{-1}s^{-1}$. Therefore, between the base temperature of 1.5K and room temperature, there is much less than an order of magnitude change in mobility. This contrasts greatly with GaAs/AlGaAs systems, where phonon scattering reduces mobility by many orders of magnitude. This is typical of such ternary channel material, where alloy disorder and clustering severely limit mobility at low temperature, but which has high conductivity at room temperature. Note, however that the mechanism responsible for the mobility peaking,

first discussed in section 5.3, is still present at these elevated temperatures and even exists at room temperature.

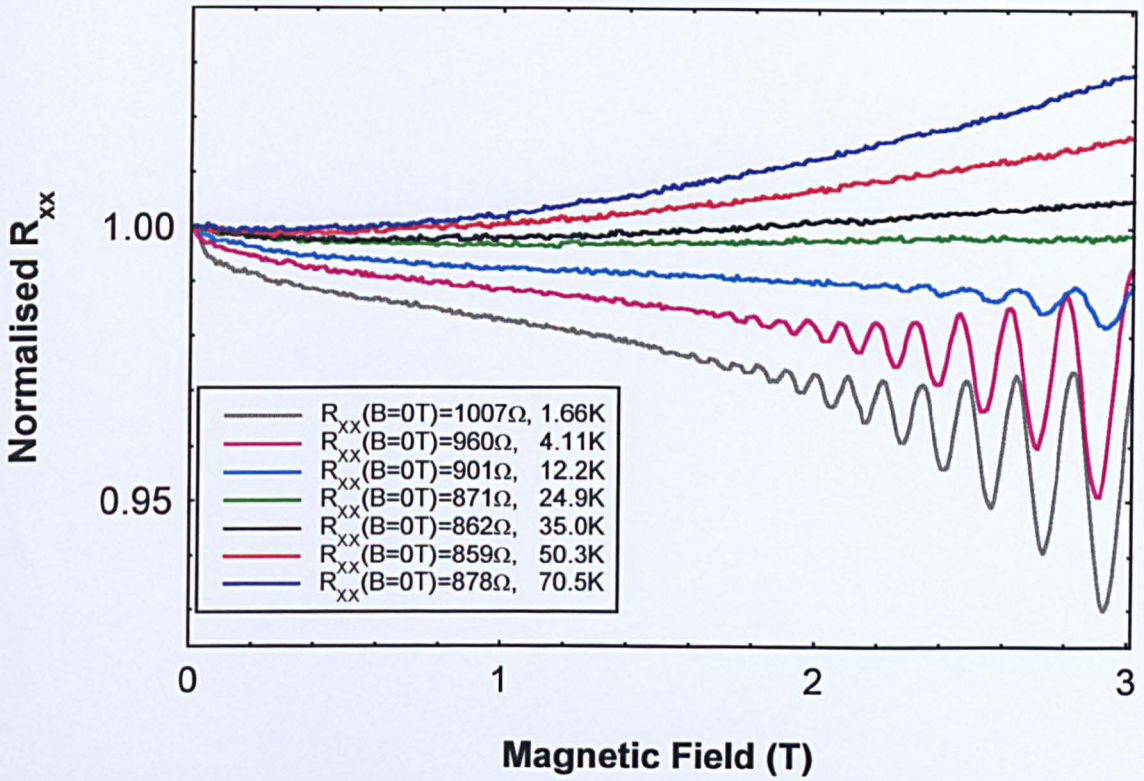


Figure 5.5.2 The normalised magneto-resistance data for sample A1236 (20% indium, 2.5nm spacer; V_g=0V) at various temperatures. Note that as temperature rises the SdH oscillations are suppressed and that a $\mathbf{B}^2$ background develops in the data above about 35K.

In order to ascertain whether parallel conduction evolved with higher temperature, a series of magneto-resistance runs were carried out. These are shown in fig. 5.5.2, where it can clearly be seen that a quadratic (B^2) trend develops with magnetic field above 35K. This is a possible signature of parallel conduction and suggests that it may play some part in limiting the mobility at the higher temperatures. Another interesting feature is the persistent negative magneto-resistance of the curves at low temperature. This is the negative magneto-resistance, which extends well beyond typical characteristic fields (<<1T) associated with weak localisation. The conductivity is also very high for the detection of weak

localisation effects. More on weak localisation will be discussed in section 5.9. The reasons for this negative magneto-resistance are unknown. Curiously, the temperature at the changeover of the sign of the gradient is approximately that where the mobility reverses its upward trend with decreasing temperature in fig. 5.5.1.

b) At low carrier density, the mobility-carrier density relation in fig. 5.1.1 is approximately linear thus extending to higher temperatures the suggestion made at low temperature in section 5.3 of dominant long-range cluster scattering intrinsic to the channel. From the zero mobility extrapolations it suggests that there is a progressive change in the carrier density trapped in the potential fluctuations of the channel. As the temperature rises electron trapping tends to zero. In other words, as the thermal energy increases the electrons are activated out of these low-temperature confining potentials. Closely connected with this behaviour at slightly higher density (about $5x10^{11}cm^{-2}$), the general reduction of mobility, or more correctly conductivity, with decreasing temperature suggests behaviour reminiscent of weak localisation. The magnitude of the changes in conductivity with temperature is typical of this mechanism. With decreasing temperature, the average length between phase-destroying collisions - the phase coherence length (l_ϕ) - greatly increases relative to the elastic mean free path (l). Therefore, the probability of coherent back scattering increases when temperature and carrier density are reduced with a reduction in inelastic scattering from phonons. However, it is likely that potential fluctuations arising from the donors and the indium clustering will confuse the issue somewhat. Separate conductivity measurements made at high carrier depletion confirmed this characteristic in some samples.

More will be discussed later in section 5.9, as will the presence or otherwise of weak localisation using magneto-resistance techniques.

c) As the temperature is reduced below 100K, the peaking feature and its cause becomes relatively more defined as the influence of the inelastic phonon scattering declines. A similar behaviour was apparent in section 5.4, where the relative contributions of RII scattering changed in response to an applied bias during cooling. At high density in fig. 5.4.2, there is some qualitative similarity in the data for sample A1046 (and A1190) under various bias cool voltages with that in fig. 5.5.1 in the way the peak evolves. In both cases, the peak becomes more defined with a lowering of the relevant scattering influence - RII scattering from uncorrelated donors in fig. 5.4.2 and phonon scattering in fig. 5.5.1. In the former case, this was another reason for rejecting RII scattering as the reason for the actual mobility peaking.

d) At very low carrier density there is an extension of the trend of more electrons condensing into the conduction band-edge fluctuations with decreasing temperature. This would be expected with the corresponding reduction in thermal energy well below that of the activation energy (E_A).

e) At higher density above the mobility peak, there is a further fall in mobility with temperature reduction, where weak localisation would be difficult to detect. Examination of fig. 5.5.3, which is a similar plot for the other sample with a 2.5nm spacer (A1190), also shows the same qualitative effect at and above the peak. However, here the magnitude of the fall with decreasing temperature is smaller, but the mobility reversal occurs at much the same temperature. Fig. 5.5.4 shows a similar set of measurements for a sample with a 5nm spacer (A1235),

with no fall in mobility apparent at lower temperatures for higher carrier densities. The electrostatics existing around this thicker spacer sample prevents the creation of a higher channel density. Samples with a 10nm spacer displayed much the same features as the 5nm specimen.

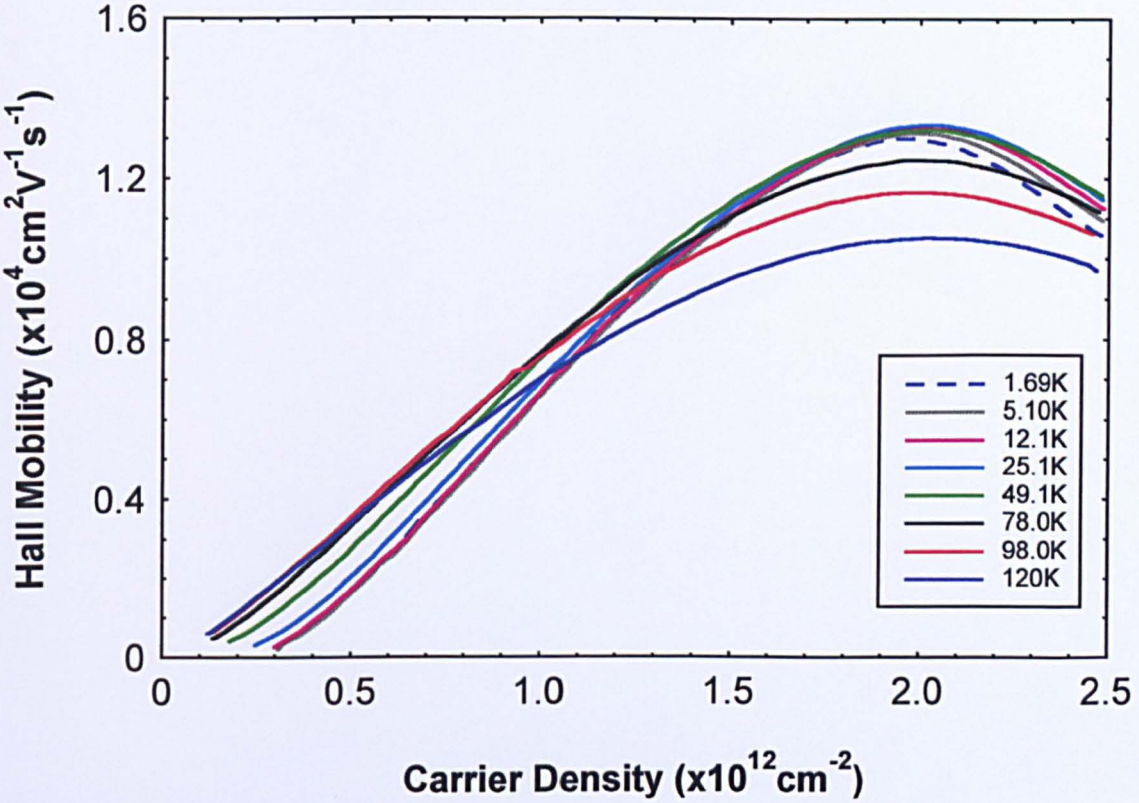


Figure 5.5.3 The variation of Hall mobility with carrier density at variable temperature after a normal cool for sample A1190 (30% indium, 2.5nm spacer). The top temperature was restricted to avoid complications with *DX* centre occupancy.

All the evidence suggests that the anomalous reversal of the increase of mobility with decreasing temperature is mainly dependent on the spacer-width and the channel-spacer barrier energy offset. The effect is strongest when the spacer width and barrier energy are both minimised (A1236). It is therefore tempting to believe that the effect has a strong association with factors in the doping layer, just as hypothesised for the peak characteristic itself. As the indium is known to segregate and 'float' upwards with the MBE growth, the possibility

exists that the wavefunction tail is interacting with this indium or its resultant effects at the interface of the spacer and the doping layer. However, it is very likely that the highly strained 30% indium sample (A1190) would show a stronger peaking here than the similar 20% indium sample (A1236). This is not observed. Suggesting also that the indium aggregating at the spacer-doping layer interface is responsible for the mobility peak is also not supported by observation. As the

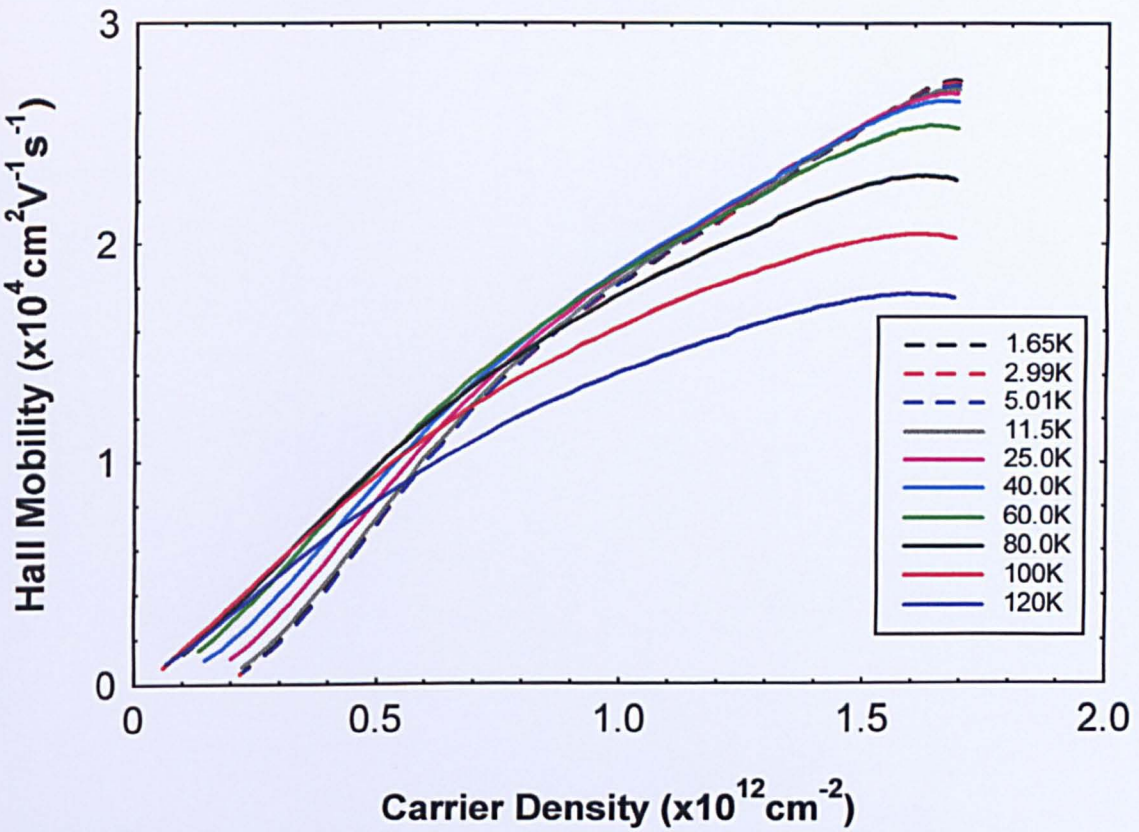


Figure 5.5.4 The variation of Hall mobility with carrier density at variable temperature after a normal cool for sample A1235 (20% indium, 5nm spacer). Note that on this occasion there is no fall in mobility at high carrier density as the temperature falls.

growth temperature was the same for both the samples, A1236 and A1190, then the 30% sample (A1190) would have the rougher interface and hence the stronger peak. Again, this is not observed.

Another interesting observation is that in both data sets containing peaks, there is a tendency for a point of inflexion to evolve in the curves at high carrier density well above the peak. This is more apparent at lower temperature, but the reason for this is unknown. Whatever the correct explanations for these characteristics, it is notable that for **all** densities the mobility at 80K is higher than that at 1.66K.

5.5.3 Section Summary

The observation of a resistance minimum in the sample with a 20% indium content and a 2.5nm spacer as it cooled to base temperature prompted a series of variable-temperature measurements. The two main parameters studied under these conditions (about 1.5K to 130K) in the metallic regime were resistance and mobility. The upper temperature was limited to avoid any complications arising from thermal activation of the *DX* centres. All data was acquired 4-terminally with the mobility calculated using the classical Hall technique. This method is prone to be inaccurate at higher temperatures where additional conducting pathways can be thermally activated. Magneto-resistance plots for a range of samples suggested the onset of possible parallel conduction at a temperature above about 30K. A reduced weighted mobility, particularly at high carrier densities, was the result. At the highest temperature, the onset of phonon scattering lowered the mobility further, and the mobility peak, where applicable, becomes slightly less well defined. At high densities in samples with the thinnest spacers there was an additional reduction in transport mobility as the temperature fell below 30K to the base temperature. This is of course related to the initial observation made above concerning resistance during cooling. The exact

reason for this remains unclear but it does not exist in samples with thicker spacers.

At lower carrier densities, all samples tested showed a general decrease in mobility or conductivity with decreasing temperature. The magnitudes of the changes are reminiscent of weak localisation and the onset of coherent back scattering. This weak localisation behavour is a precursor to evidence of a temperature dependent strong localisation. In the data for all samples, the zero mobility cut-off shifts to lower carrier density with increasing temperature. This suggests that trapped electrons within the potential fluctuations at high depletion can be thermally activated out of these as the temperature is increased.

5.6 Mobility Anisotropy at Low Temperatures

Six samples were tested for mobility anisotropy at base temperature. This was done to ascertain if the assumed indium clustering was isotropic in distribution. They consisted of six samples - half of which were grown at a nominal 520°C and the remaining at 500°C. Of the former, two had 20% indium channels (A1236 and A1235)[‡] and the other a 30% channel (A1242). All of the latter three are sub-critically doped specimens - all with a 30% indium channel. These three along with A1242 were also tested for anisotropy at room temperature, but these results will be discussed along with the other room temperature results in section 5.7.

The data is presented in figs. 5.6.1 to 5.6.3 below. The Hall mobility versus carrier density for all of the 30% indium channel samples is plotted in fig. 5.6.1. The corresponding plots for the two 20% indium samples are shown in figs. 5.6.2 and 5.6.3. For the latter two samples, the mobility is measured in each of 4 directions [011], [01 $\bar{1}$], [001] and [010]. The other four samples have only three of these directions tested per sample. There is, particularly at higher carrier density, a clear anisotropy in all samples tested, with the [01 $\bar{1}$] direction **always** having the highest mobility and the [011] direction **always** having the lowest. The two intermediate directions ([001] and [010]) **always** have an intermediate mobility, and for samples where both intermediate directions were measured, the values were in good agreement.

[‡] I am grateful to Dr. C.J. Emeleus, Mr. A. Watt and Mr. M. O'Leary for supplying the data for these 2 samples.

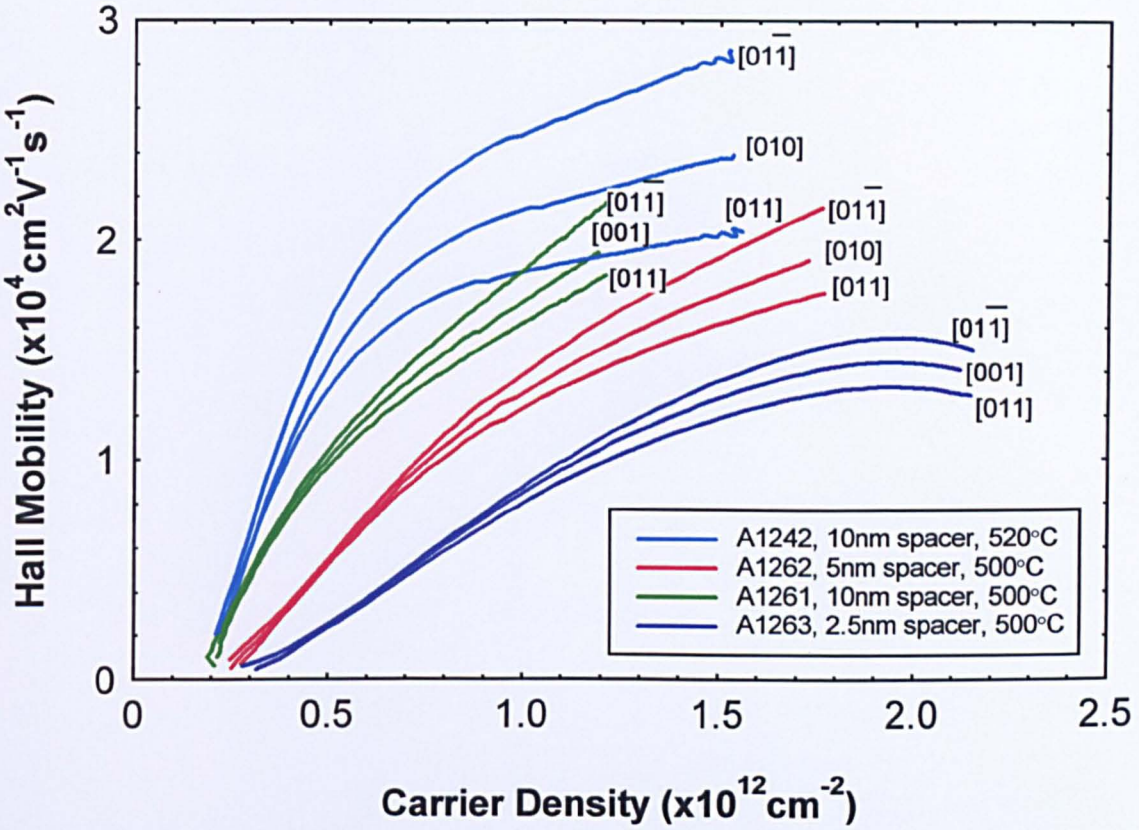


Figure 5.6.1 The mobility-carrier density plots, acquired at 1.5K, for all the measured 30% indium channel samples showing clearly the mobility anisotropy. Note that the sample grown at 520°C has the largest anisotropy.

It should be emphasised at this point that the substrate supplied was specified as nominally flat to ±0.1°, therefore substrate misorientation was rejected as a cause of the anisotropy. It is notable that the greatest anisotropy difference occurs for the sample with the highest indium content and growth temperature (A1242). It is also apparent from figs. 5.6.1, 5.6.2 and 5.6.3 that the samples with the high indium content (30%) coupled with the lower growth temperature (500°C) result in much the same anisotropy as those with the lower indium content (20%) coupled with the higher growth temperature (520°C). These observations are intuitively in agreement with the indium segregating

and clustering to a greater or lesser extent depending on the temperature and the indium content. A larger temperature and indium content, perhaps also driven by the inherent strain, would tend to segregate and cluster more strongly than examples with low indium content and growth temperature.

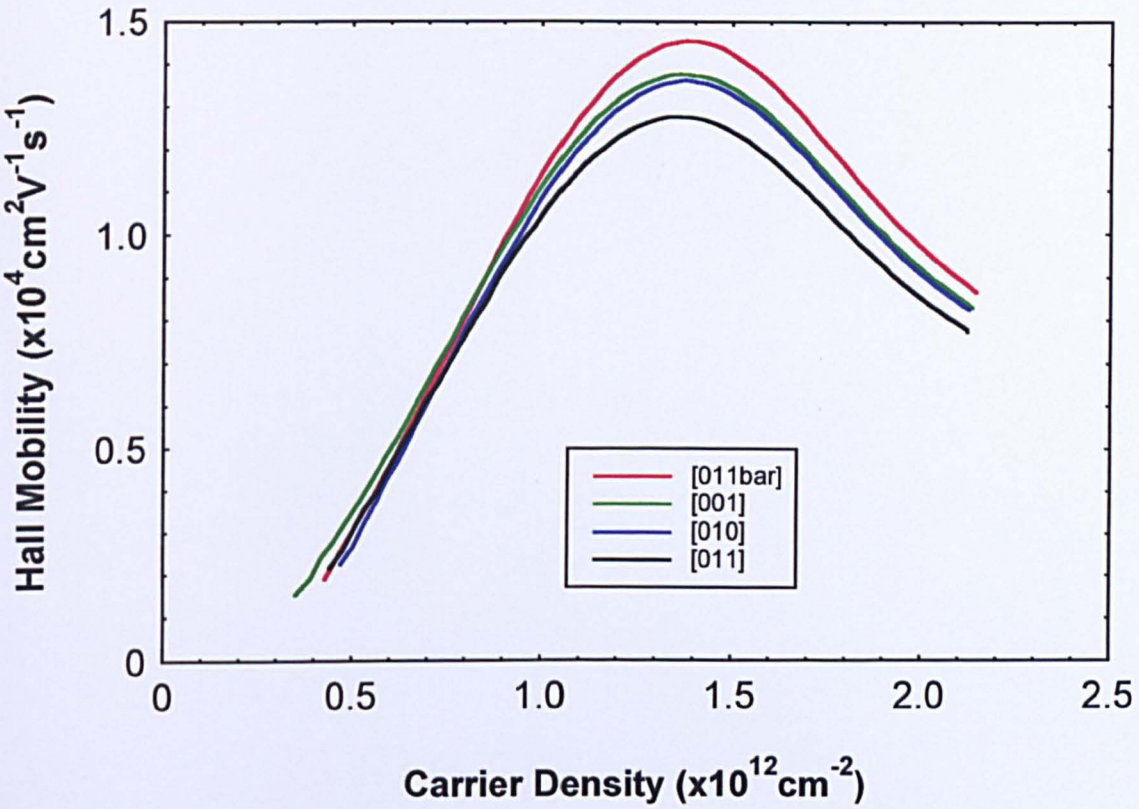


Figure 5.6.2 The mobility-carrier density plots for sample A1236 (2.5nm spacer) which has a 20% channel indium content grown at 520°C. The data was acquired at 4 different directions at 1.5K.

Why this effect is directional is more difficult to reason. RHEED observations‡ made during MBE growth, suggest that on "arsenic-stable" (100) surfaces, reconstruction along step-edges is orthogonally anisotropic. Thus the combination of the indium segregation and the two different step-edge densities in orthogonal directions may explain the mobility anisotropy. Schweizer et al (1991)[6] also observed mobility

‡ Private communication – Prof. C.R. Stanley.

anisotropy in a GaAs/$In_{0.2}Ga_{0.8}As$/$Al_{0.3}Ga_{0.7}As$ inverted HEMT. They found that provided the 'critical' thickness of the quantum well was not exceeded, then a similar mobility anisotropy (at 77K) resulted as is

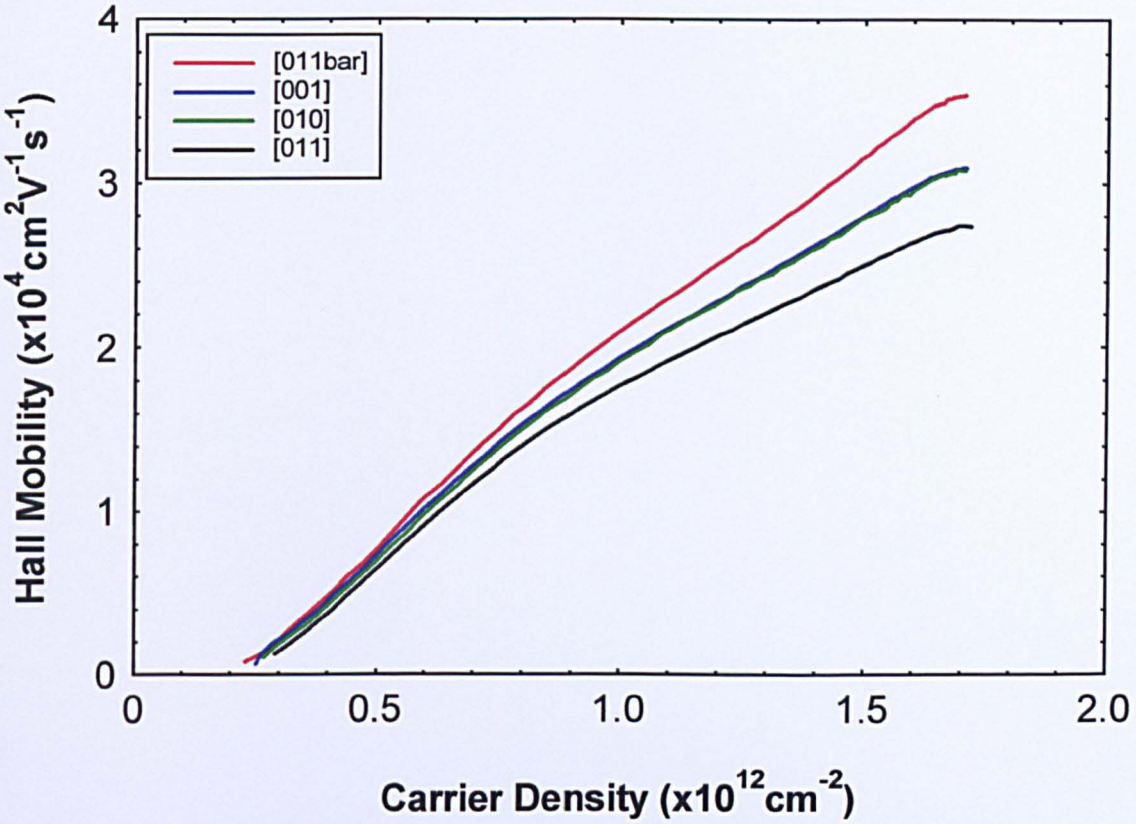


Figure 5.6.3 The mobility-carrier density plots for sample A1235 (5nm spacer) which has a 20% channel indium content grown at 520°C. The data was acquired at 4 different directions at 1.5K.

found in this work. They assigned the anisotropy to an asymmetry in dislocation density. If the critical thickness was exceeded, then the high and low mobility directions reversed and their mobility difference increased. Beck and Ilegems (1996)[7] also measured mobility anisotropy in an indium phosphide (InP)-based pseudomorphic HEMT (InP/$In_{0.52}Al_{0.48}As$/$In_yGa_{1-y}As$ for y=0.53-0.80). They found that the mobility anisotropy increases with growth temperature, channel thickness and indium concentration in the channel.

In this work, measurements were made on the absolute and relative mobility anisotropy as it varied with carrier density. The former is generally non-linear – usually monotonically increasing with carrier density, with the sample having a mobility peak a notable exception. However, the relative mobility anisotropy is more linear.

Measurements not yet made in the strongly localised regime may reveal differences in the magnitude of activated transport with direction. Specialist microscopical measurements were also not available within the time-scale of the project. Shubnikov-de Haas plots confirmed the values of the mobility derived from the classical Hall technique and each direction also had identical carrier density depletions. Some magneto-resistance measurements were also made on one sample (A1242) to determine the magnitude of quantum mobility and its dependence on the direction. These results will be discussed in section 5.8. The zero-magnetic field resistance in the three directions $[01\bar{1}]$, [010] and [011] in sample A1242 are 446Ω, 532Ω and 617Ω respectively – there is also large anisotropy in this parameter.

5.7 Room Temperature Measurements

A set of similar measurements was also performed on a selection of samples at room temperature (~300K). This was done not only to ascertain the full effect of phonon scattering on mobility but also to determine the linearity of electron depletion. Both these factors would be of interest to device engineers or modellers. The data was acquired without the cryostat, using an electro-magnet, whose field was set at 0.7T. The remaining equipment used was much the same, but pre-amplifiers were not used. The sample carrier was mounted in a simple holder, with little noise suppression, consequently data had to be carefully acquired at times of 'quiet environmental electrical noise'.

The mobility data is presented in fig. 5.7.1, where the fall in mobility from the previous values is immediately apparent. However, the fall is not as dramatic as that for GaAs/AlGaAs systems, with their residual mobility at this temperature often lower than the values found here. The different sample curves have qualitatively much the same relative mobilities as they had at base temperature. They are slightly more compressed together in 'mobility-space', with the best mobility value for a 10nm spacer sample barely twice the worst for a sample with a 2.5nm spacer. Two other interesting observations emerge from this data. 1) There is clearly no residual carrier density in the potential fluctuations at this temperature. Extrapolation of the mobility curves to zero mobility reveals no finite carrier density cut-off. All the electrons can thus make a contribution to the metallic conductivity. 2) The mobility peaking feature, at least for one of the samples, is still strongly represented. If this feature has its origin in an elastic scattering process,

then it is quite remarkable that it persists until room temperature despite heavy masking from phonon scattering.

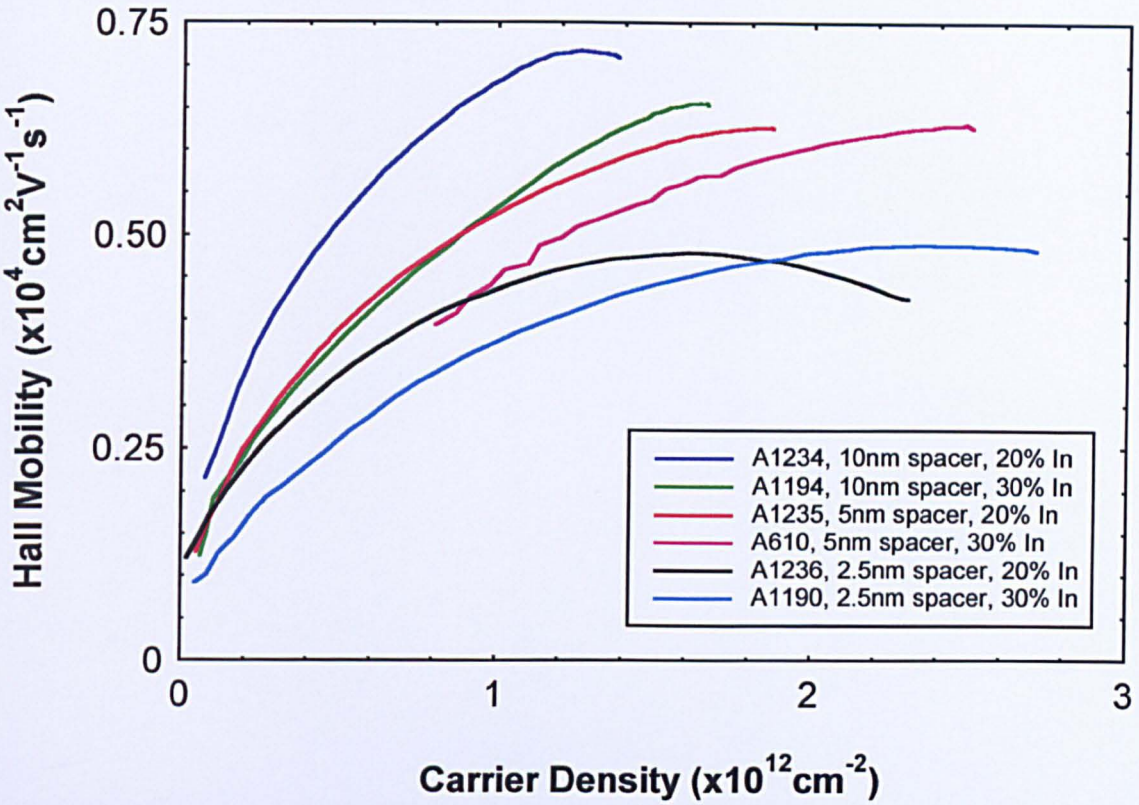


Figure 5.7.1 The mobility-carrier density plots for a selection of samples acquired using Hall techniques at room temperature (about 300K). Note the more compressed nature of the curves at this temperature when compared to those at base temperature. There is also no residual electron density trapped in the potential fluctuations within the channel. The peak of A1236 is still apparent even at this elevated temperature.

Also shown separately in fig. 5.7.2, are the mobility curves for each of the sub-critically doped samples along with that for sample A1242. Each sample has two data sets representing the high and low mobility directions as discussed previously in section 5.6 for base temperature. As in the case above with the peaking characteristic, this feature of mobility anisotropy also persists to room temperature, although not as strongly as that at base temperature. It is worth remembering that the three samples, A1261-3, although having a large

30% indium content, were grown at a nominal 500°C - lower than the other 30% samples and all the 20% samples (520°C).

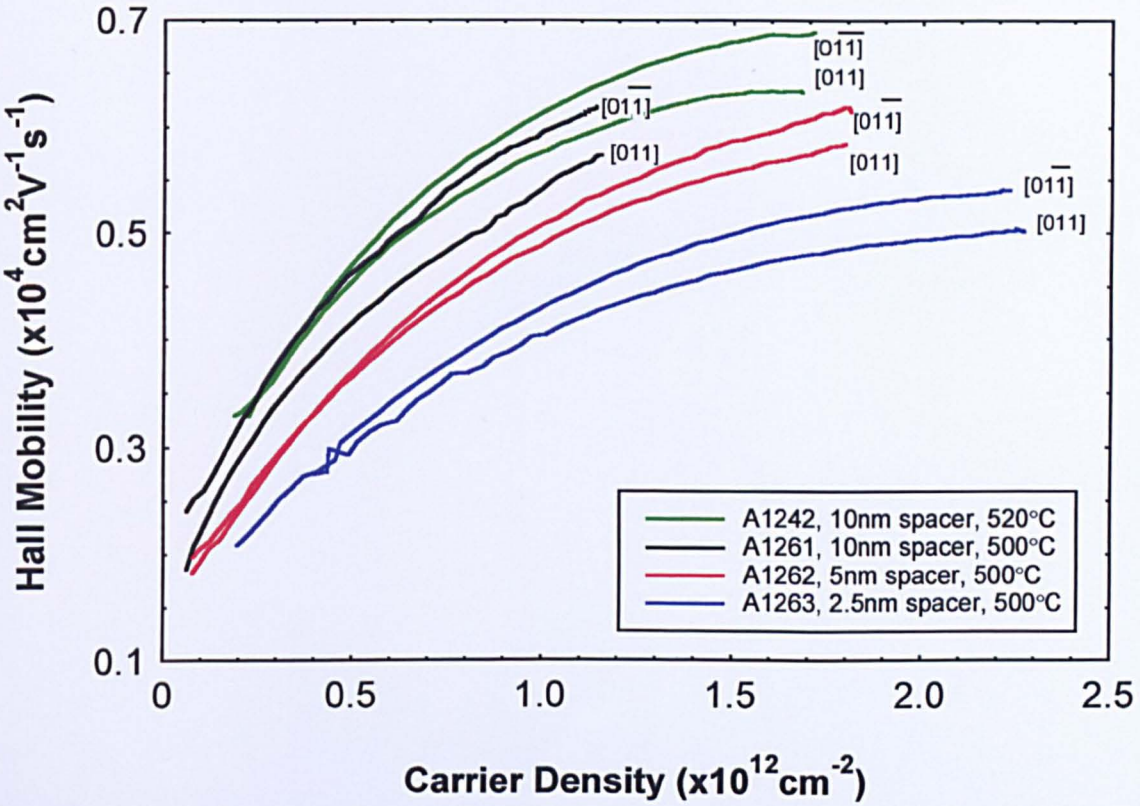


Figure 5.7.2 The mobility-carrier density plots of a selection of samples with 30% channel indium content acquired using Hall techniques at room temperature (about 300K) in each of the high and low mobility directions.

The carrier density depletion curves are shown in fig. 5.7.3. Data for the sub-critically doped samples has also been included as they were primarily grown to study the effect of this doping strategy on the depletion. For a given spacer width, the sub-critically doped specimens deplete with better linearity, with the 5nm spacer a particularly good example (A1262). If good linear depletion combined with reasonable conductivity at room temperature were a device requirement, then the sample with 30% indium, 5nm spacer and doping around the 'critical' value would be a good recommendation.

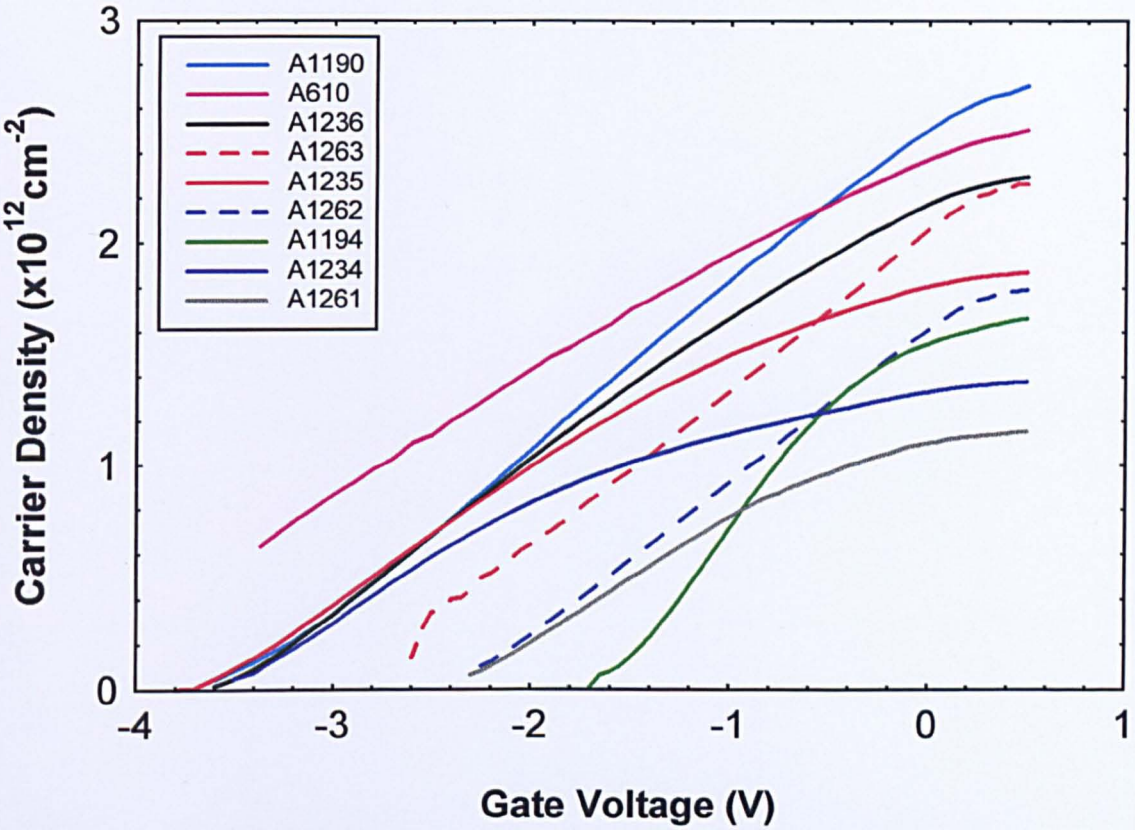


Figure 5.7.3 The carrier density depletion curves for a selection of samples acquired at room temperature (about 300K). Note that the 3 sub-critically doped samples (A1261-3) have the best linearity, with the 5nm spacer specimen (A1262) particularly good. The presence of over-doping in the other samples particularly A1194 and A1234 promotes poor linear depletion.

5.8 Studies of Quantum Mobility

5.8.1 Introduction

Until now all of the discussion concerning the electron mobility within these layers has centred on the classically derived transport mobility (μ_{tr}). The variations of this parameter have offered many insights, but more information can be obtained if the quantum mobility (μ_q) is also extracted. The main difference between the two mobility types is in the emphasis given to scattering angle – that is, the angle of the $\boldsymbol{k}+\boldsymbol{q}$ vector relative to the initial $\boldsymbol{k}$ vector in momentum space, see fig. 2.4.1. Transport mobility is weighted towards large-angle scattering, while quantum mobility treats all angles equally. Equation 2.4.30 derived for RII scattering is strictly related to the transport time and hence transport mobility, but the quantum or single-particle scattering time and hence quantum mobility can be obtained simply by dividing the integrand by the factor ($1-\cos\theta = |\mathbf{q}|^2/2|\mathbf{k}|^2$), which incorporates the appropriate angular adjustment. A charged scattering centre sited remotely from the electrons would tend to scatter by smaller angles than one sited intrinsically within the channel. Hence, if both sources of scattering are present with the channel scattering dominant, it will produce quantum and transport mobilities closer to one another in value when compared with dominant RII scattering.

Experimentally, the quantum mobility is extracted from the quantum relaxation time (τ_q) in equation 2.2.2 using the relation $\mu_q = e\tau_q/m^*$ or an equivalent. A 'Dingle plot', described by the expressions in 2.2.7 and 2.2.8, characterises the decay of the exponential envelope of the

oscillations with decreasing magnetic field. It will have a gradient from which the quantum mobility can be obtained. Higher fields may require a modification using expression 2.2.11. For analysis, it is more useful to express the quantum mobility as a ratio with the transport mobility (μ_{tr}/μ_q), as both change with carrier density. Measurements were acquired 4-terminally at 1.5K, as at higher temperatures the SdH oscillations become suppressed. As these are low-mobility samples, the temperature correction in the mobility formula is always less than 2% and will henceforth be ignored. The effects of bias cooling and crystal direction on the mobility ratio were tested and the behaviour around a transport mobility peak for a 2.5nm spacer sample was also analysed.

Before proceeding, a number of adjustments were made to the relevant equations presented in section 2.2, equations 2.2.(2-9), for experimental ease. The peak to peak (Pk-Pk) amplitudes of the magneto-resistance oscillations were measured and normalised with respect to the zero-field resistance multiplied by a factor of eight – double the normal factor of four (equation 2.2.2) used when amplitude is measured. In order to obtain the gradient related to the actual quantum mobility, the natural logarithm of this normalised quantity was calculated. As the data was normalised and 'logged', the expected extrapolated value on a Dingle plot at infinite magnetic field is therefore zero. The quantum mobility (μ_q) is extracted from a re-expression of the exponent term (now denoted as A) in equation 2.2.6 in the following way:

$$A = \frac{-\pi}{\omega_c \tau_q} = \frac{-\pi}{\mu_q B} \qquad (5.8.1)$$

where ω_c is the cyclotron angular frequency, and is expressed by $\omega_c = eB/m^*$, with $\tau_q = \mu_q m^*/e$ and both these expressions have been substituted into the middle term to obtain the right-hand term. If the

numerator and denominator of A are both multiplied by the filling factor (ν), then the appropriate gradient (P) of a Dingle plot emerges as:

$$P = \frac{-\pi}{[\nu B]\mu_q} \qquad 5.8.2$$

if the independent variable is ν. (The filling factor (ν) is simply a convenient parameter to use as the abscissa of the plot and it simplifies the experimental determination of the quantum mobility. It is related to the way the states at the Fermi level are filled as the magnetic field is varied.) Therefore the quantum mobility is:

$$\mu_q = \frac{-\pi}{[\nu B]P} \qquad 5.8.3$$

The product [νB] is constant for a given SdH curve.

5.8.2 Quantum Mobility Observations

In this sub-section, the behaviour of quantum mobility in all the conditions described in the introduction above will be discussed. In all samples tested at a given bias cool voltage, there was always a general rise in quantum mobility with increasing gate voltage. A typical Dingle plot is shown in fig. 5.8.1 for sample A1194, see also section 2.2.2. This rise in quantum mobility with gate voltage is primarily due to the increasing screening by electrons entering the donor layer, but there is also some self-screening by the channel electrons themselves. It was also found, in general, that the curves for gate voltages around zero extrapolated approximately to zero on the ordinate axis of a logarithmic Dingle plot. As gate voltage becomes more negative, there is a greater likelihood for the corresponding lines to extrapolate above zero. The

lines referred to are 'best-fit' straight lines and on occasion disguise any curvature of actual data – either towards or away from a zero extrapolation. No rigorous theory yet exists which can account for this displacement, but it can be shown‡ that provided the data does not appreciably curve, then it can be interpreted as a genuine quantum

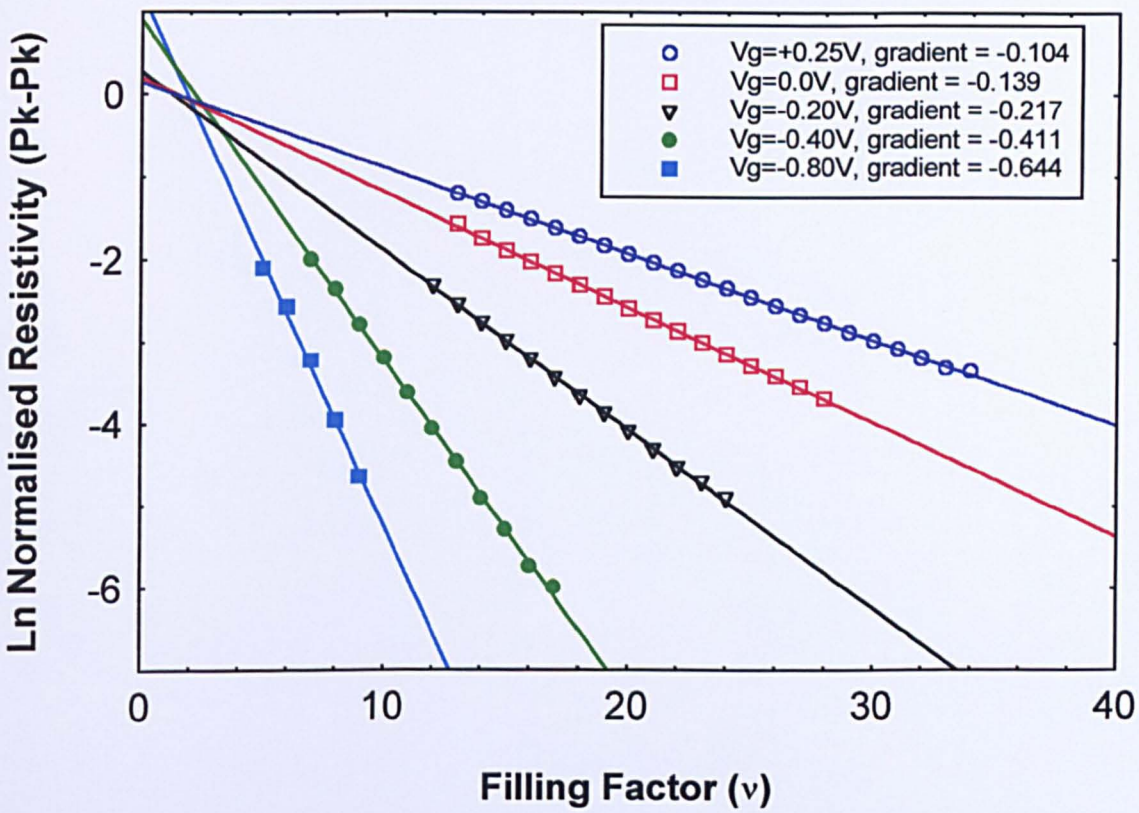


Figure 5.8.1 The Dingle plot acquired at 1.5K for sample A1194 (10nm spacer, 30% indium) after a normal cool. Note the tendency for data acquired at more negative gate voltage to extrapolate to above zero on the ordinate axis.

mobility despite its failure to extrapolate to zero. The difference can simply be regarded as an offset. Data which curves and cannot be explained by other reasons such as high-field effects must be regarded with scepticism and this will be mentioned at the appropriate place. A comparison of data sets dependent only on gate voltage can be misleading, as it does not account for differences in carrier density

‡ P.T. Coleridge Private communication.

which, as stated above, can provide extra screening from scattering by charged centres. Therefore most of the remaining analysis will relate a given parameter to carrier density, thus reducing hidden variables.

Before that however, a typical Dingle plot showing the variation of the gradient with the bias applied during cooling is shown for sample A1194 in fig. 5.8.2. The data was acquired at a temperature of 1.5K using a gate bias of zero. All the other tested samples were qualitatively

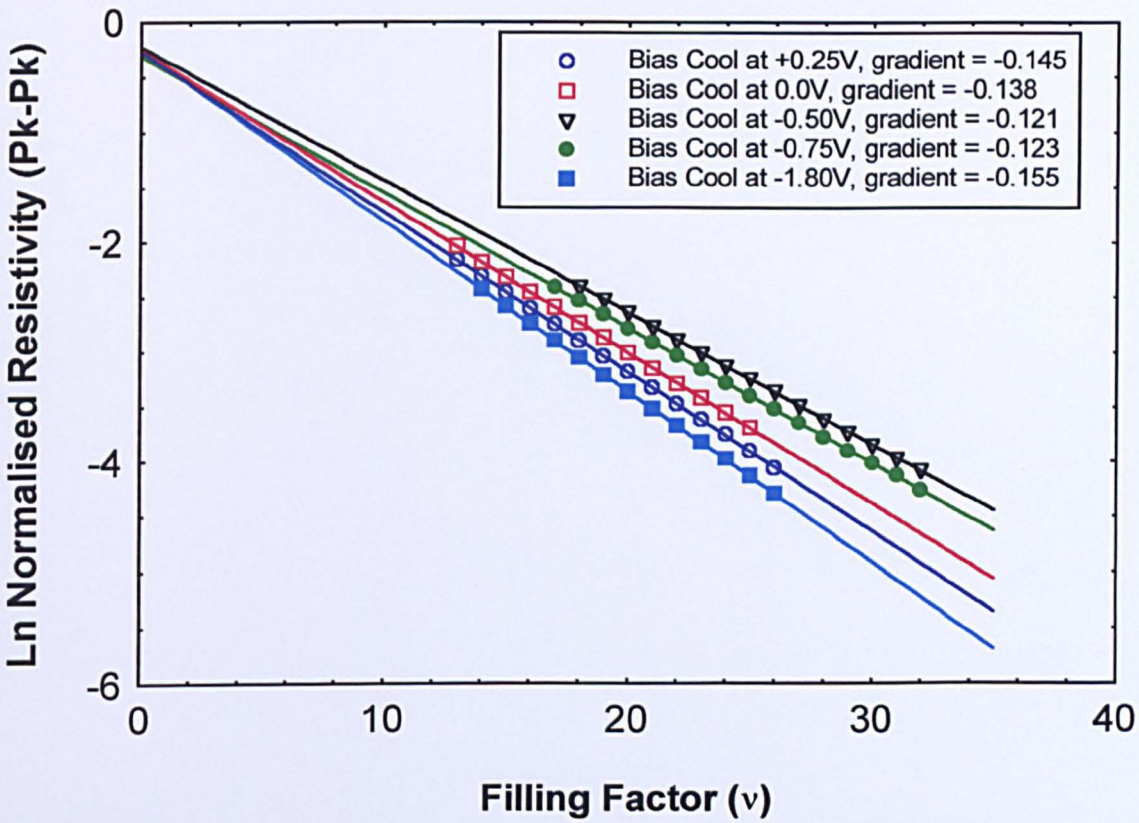


Figure 5.8.2 The Dingle plot acquired at 1.5K for sample A1194 (10nm spacer, 30% indium) as the bias cool voltage was varied. The gate voltage was set at 0V. Note the reversal in the gradient magnitude as the bias cool voltage is decreased.

similar only varying in their relative gradients and extrapolated values at zero filling factor. The example shown extrapolates to just below zero. This effect can, in principle, be associated with weak localisation-like behaviour, but the conductivity is quite high in this instance.

In fig. 5.8.3 is shown the variation of quantum mobility (μ_q) with both carrier density and bias cooling for four samples (A1046, A1047, A1190 and A1194). Note that samples with thicker spacers (A1194: 10nm and A1047: 5nm) have both a much higher quantum mobility at

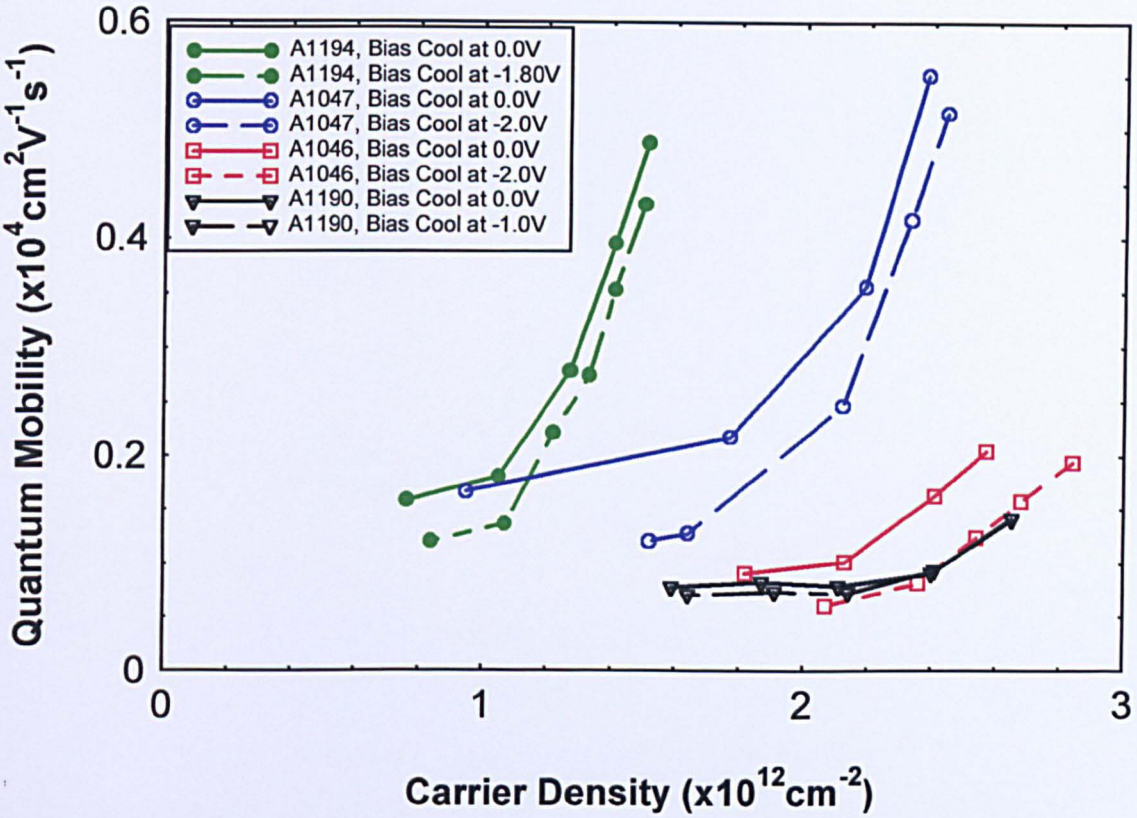


Figure 5.8.3 The variation of quantum mobility with both the carrier density and the bias cool voltage at 1.5K, for four 30% indium samples (A1194: 10nm spacer; A1047: 5nm spacer; A1046 and A1190: 2.5nm spacer). Note the higher quantum mobility in samples with thicker spacers – due to not only this thickness but also the degree of overdoping.

high carrier density and a much greater change in quantum mobility over the measured carrier density range. The high quantum mobility is probably due to screening by electrons in the doping layer – the thicker spacer samples were more overdoped than the thinner spacer specimens. As these electrons are removed, there is a relatively sudden fall in the quantum mobility as this screening is removed. In contrast, the thin spacer samples (A1190 and A1046 – both 2.5nm) have both relatively

low quantum mobility at high carrier density and a small change in this mobility over the measured carrier density range. These samples, being only lightly overdoped, lack much of the screening provided by doping layer electrons but which is available to the thicker spacer samples. In addition, a thin spacer does not offer the channel electrons much shielding from long-range, small q scattering by the ionised donors.

At a given carrier density, negative bias cooling has the effect of lowering the quantum mobility of a given sample. (Due to pressure of time, A1190 was not subjected to a full negative bias cool – one in which the channel was effectively depleted during cooling.) After a negative bias cool, the *DX* centres of the doping layer are largely unoccupied. In this condition, they cannot produce the appropriate correlation by occupation. Therefore one would expect the long-range small q scattering by the ionised donors to be stronger in this case and hence reduce quantum mobility. It is also apparent, that once electrons have been removed from the shallow states of the donor layer, the effect becomes relatively stronger – a greater percentage change in the quantum mobility.

The relation between the quantum and Hall-transport mobilities for the same four samples is shown in fig. 5.8.4. In this way the relative changes with carrier density of the two mobility types can be seen, although the ratio of the two mobilities (μ_{tr}/μ_q) will be displayed in a later figure. Of the two mobilities, the quantum mobility is more likely to fall quickly with decreasing carrier density, and it has a much greater tendency to almost saturate at lower carrier densities. In addition to reasons stated above, this may indicate that as the Fermi level in the channel sinks towards the potential fluctuations of the conduction band-edge, mobility is relatively more affected by shorter-range large q scattering, despite small q scattering being at its strongest.

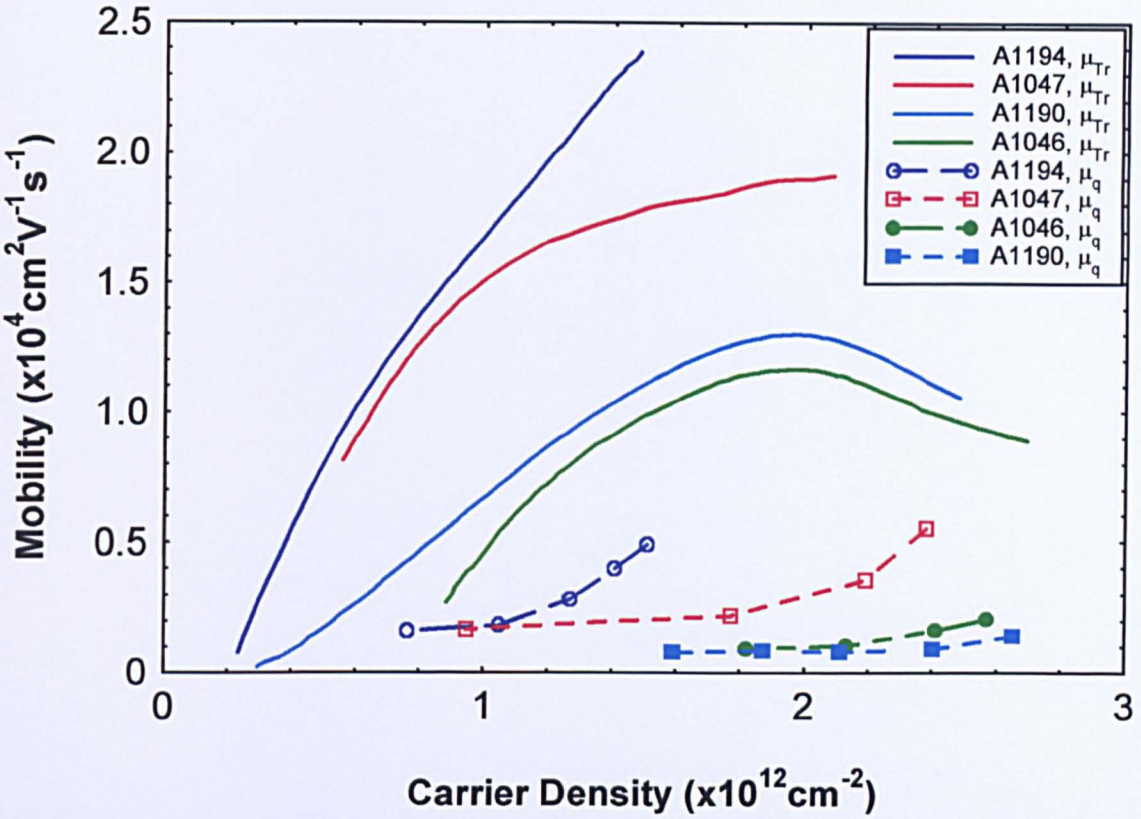


Figure 5.8.4 The relative variation of Hall-transport and quantum mobility with carrier density, acquired at 1.5K and applying zero gate voltage. The same four samples were measured as in fig. 5.8.3. Note the quantum mobility (μ_q) is more likely to almost saturate at low carrier densities when compared with the transport mobility (μ_{Tr}).

The ratio of transport to quantum mobility as it varies with carrier density and bias cooling is shown in fig. 5.8.5 for the same four samples. The data for A1047, in addition, includes bias cooling at +0.25V. The large fall in quantum mobility at high carrier density for all but the thinnest spacer samples is largely responsible for the rapid rise in the mobility ratio. However, a turning over in the data appears to evolve. There may be some doubt as to the validity of the points at the smallest carrier density of the black, red and green curves here, due to the problems of 'curving data' in the Dingle plot. It is difficult to conclude

that the peaking feature is genuine, because it is possible that the quantum mobility will continue to fall at carrier densities lower than

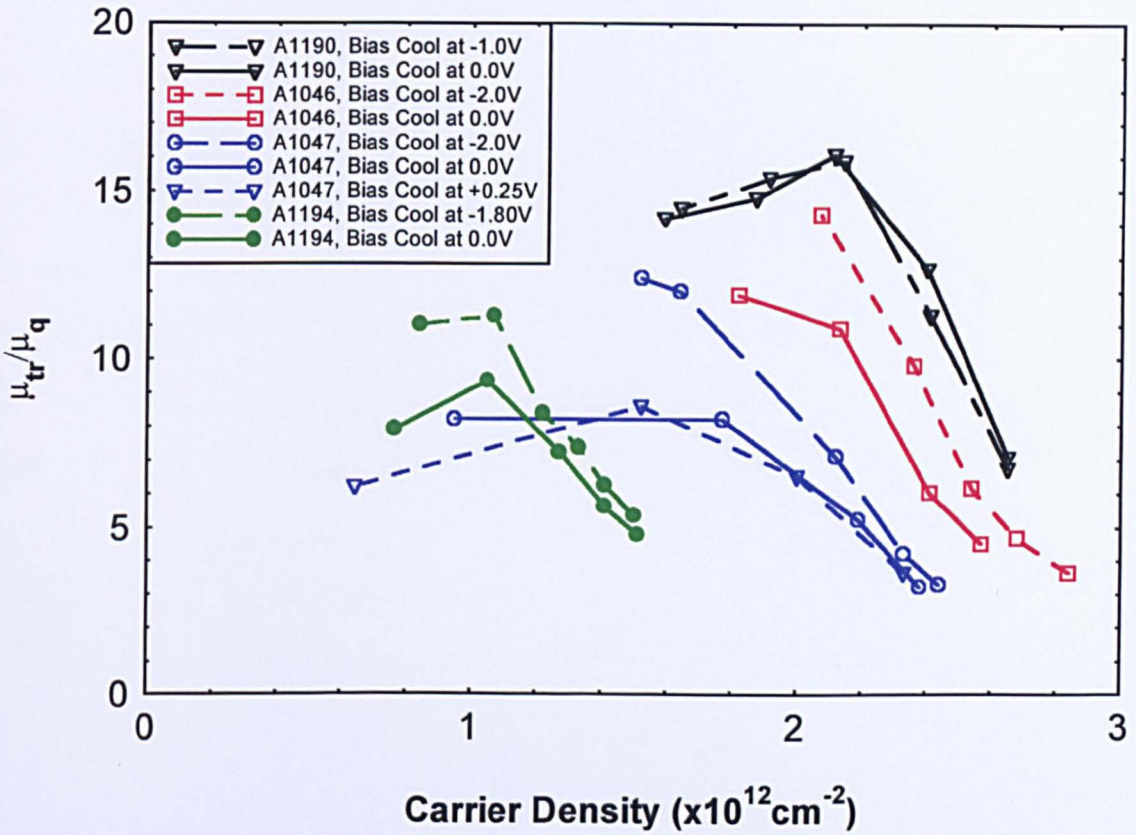


Figure 5.8.5 The relation of the ratio of transport to quantum mobility (μ_{tr}/μ_q) with carrier density and bias cool voltage for the same four samples as in fig. 5.8.3. Sample A1047 includes a positive bias cool voltage (+0.25V).

those measured. Hence, a high ratio of transport to quantum mobility could be maintained, despite the large measured fall in the former over the same low densities. Generally, for a given sample, the mobility ratio for a negative bias cool voltage rises at a greater rate than that for zero or positive bias cool voltages. This is particularly apparent at lower carrier densities, when screening electrons have been removed from the doping layer. After negative bias cooling, the full scattering effect of the uncorrelated ionised donors is experienced by the channel electrons, thus lowering quantum mobility in particular. The ratio of the two

mobilities will hence rise faster when compared with the mobility ratio rise with correlated donors after cooling under a positive or zero bias voltage.

The effect of direction on the transport mobility is substantial and is well-documented in section 5.6 on mobility anisotropy. Therefore, the effect of direction on quantum mobility would also seem to be a logical choice for study. Pressure of time limited the anisotropic study to only two samples - A1235 and A1242, the latter receiving the greater attention because of its large transport mobility anisotropy. For A1242, magneto-resistance (SdH) data was acquired in each of the three major directions [011], [010] and [01$\bar{1}$], where transport mobility was found to differ. To vary the carrier density, two different gate voltages were applied (0 and -1.2V). Sample A1235 was measured for confirmation, with only two directions studied. Sample A1236, while possessing significant anisotropy, was not tested in this manner for quantum mobility. It was restricted to a series of magneto-resistance acquisitions, which covered the region beyond the transport mobility peak.

The SdH data acquired for the three directions in sample A1242 with zero gate voltage applied is shown in figure 5.8.6. There is a large zero-field resistance change (maximum 38%) associated with the three directions. This is very similar to the corresponding percentage difference in the transport mobility found earlier at the same gate voltage. The direction having the lowest resistance corresponds to that with the highest transport mobility. The relative phase of all three curves remains constant indicating that the carrier density is similar in all three directions and this was confirmed using Hall techniques; less than a 1%

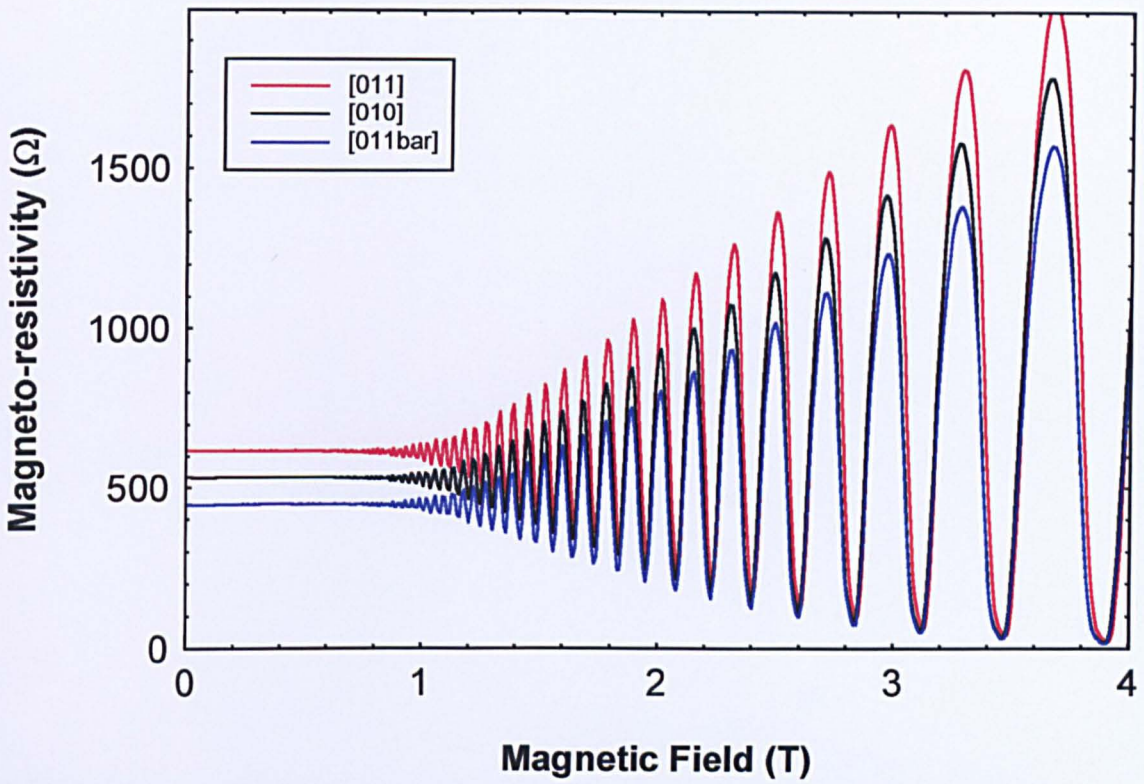


Figure 5.8.6 Magneto-resistance data acquired from Hallbars fabricated in the three denoted directions for sample A1242 (10nm spacer, 30% indium). Zero gate voltage was applied at 1.5K. Note the large zero-field resistance anisotropy (maximum 38%) between the three plots, and also that the oscillations remain in phase as the magnetic field is increased, implying no change in carrier density.

difference was observed. The equivalent set of curves acquired at a gate voltage of −1.2V is shown in fig. 5.8.7. Much the same qualitative features are found here as in fig. 5.8.6 only varying in magnitude. The respective Dingle plots are shown in figs. 5.8.8 and 5.8.9. The curves shown are a measure of the decay of the SdH oscillations with decreasing magnetic field.

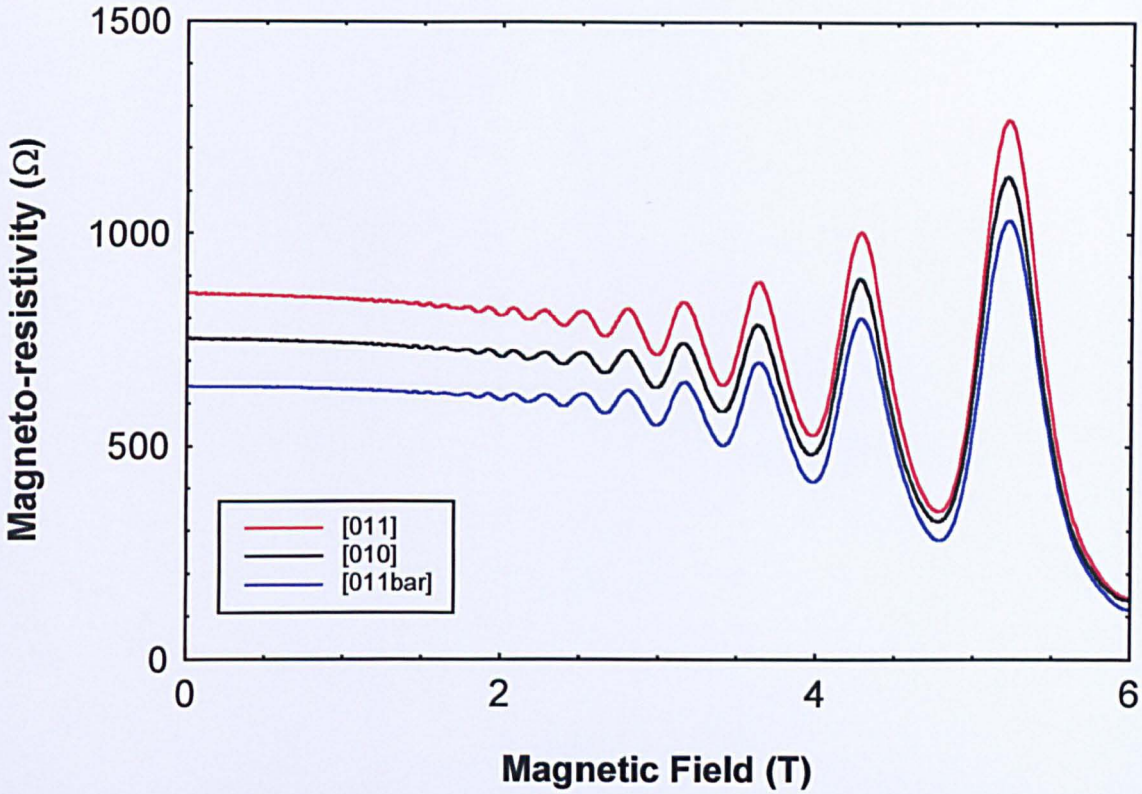


Figure 5.8.7 Magneto-resistance data acquired from Hallbars fabricated in the three denoted directions for sample A1242 (10nm spacer, 30% indium). A gate voltage of −1.2V was applied at 1.5K. Note the large zero-field resistance anisotropy (maximum 34%) between the three plots, and that as before, the oscillations remain in phase as the magnetic field is increased.

The three plots shown in each of figs. 5.8.6 and 5.8.7 have very similar νB products (less than a 0.4% difference), therefore the corresponding gradients in the Dingle plots of figs. 5.8.8 and 5.8.9 can be used directly to compare quantum mobility. It is clear from both these plots that, unlike transport mobility, quantum mobility is independent of direction. The factors that generally affect quantum mobility are normally amplified at lower carrier densities, (where electron screening in the doping layer is reduced) but no quantum mobility anisotropy was found here either. These results in combination with the transport mobility and magneto-resistance observations suggest

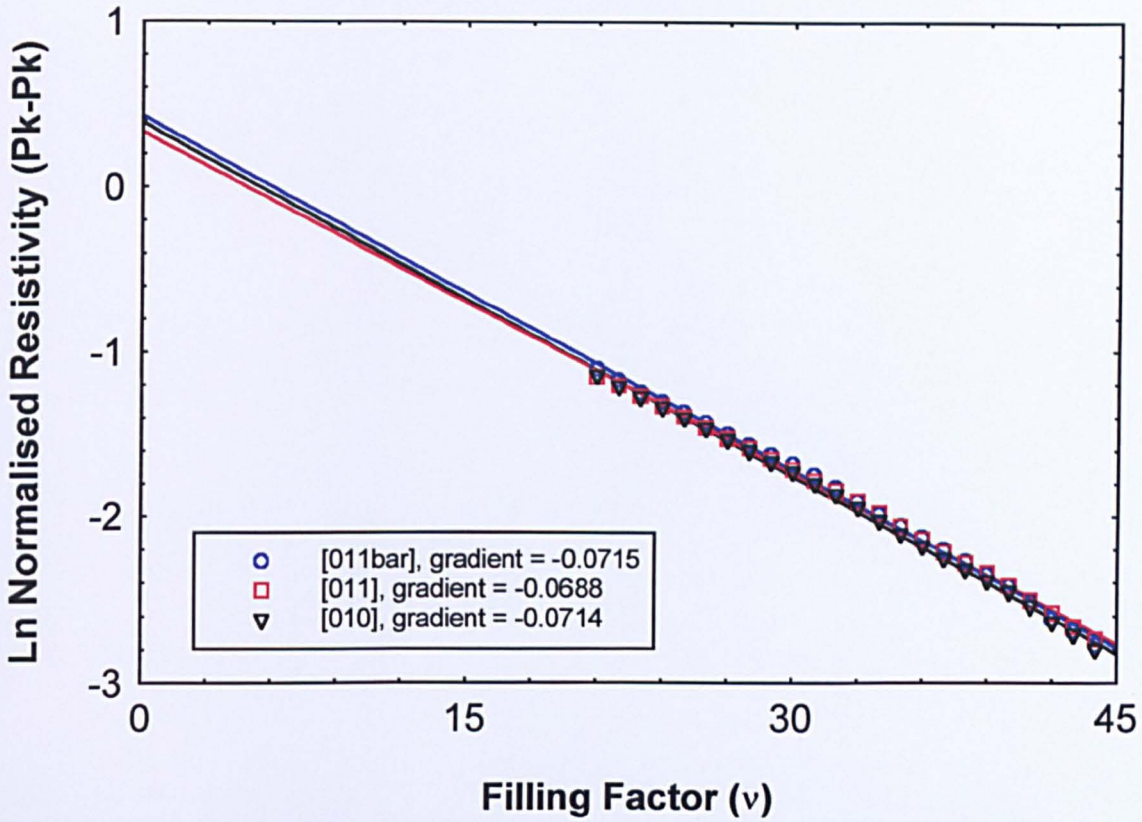


Figure 5.8.8 The Dingle plot relating to the three curves in fig. 5.8.6. Note the similarity of the gradients and of hence the quantum mobility as their νB products varies by less than 0.4% (see text). 'Pk-Pk' is peak-peak.

that the anisotropy arises from a mechanism which favours shorter-range, large q scattering and which varies quantitatively with direction. The values for quantum mobility emerging from this sample are $(0.72\pm0.01)\text{x}10^4\text{cm}^2\text{V}^{-1}\text{s}^{-1}$ at zero gate voltage and $(0.21\pm0.01)\text{x}10^4\text{cm}^2\text{V}^{-1}\text{s}^{-1}$ when the gate voltage was –1.2V. Therefore, for the directions $[01\bar{1}]$ [011] respectively, the ratios of transport to quantum mobility emerge as 3.9 and 2.8 at zero gate bias and 12.3 and 9.1 at –1.2V gate bias. The limited data from sample A1235 in two directions only ($[01\,\bar{1}]$ and [010]) confirmed that quantum mobility is also insensitive to current direction in this sample. Its quantum

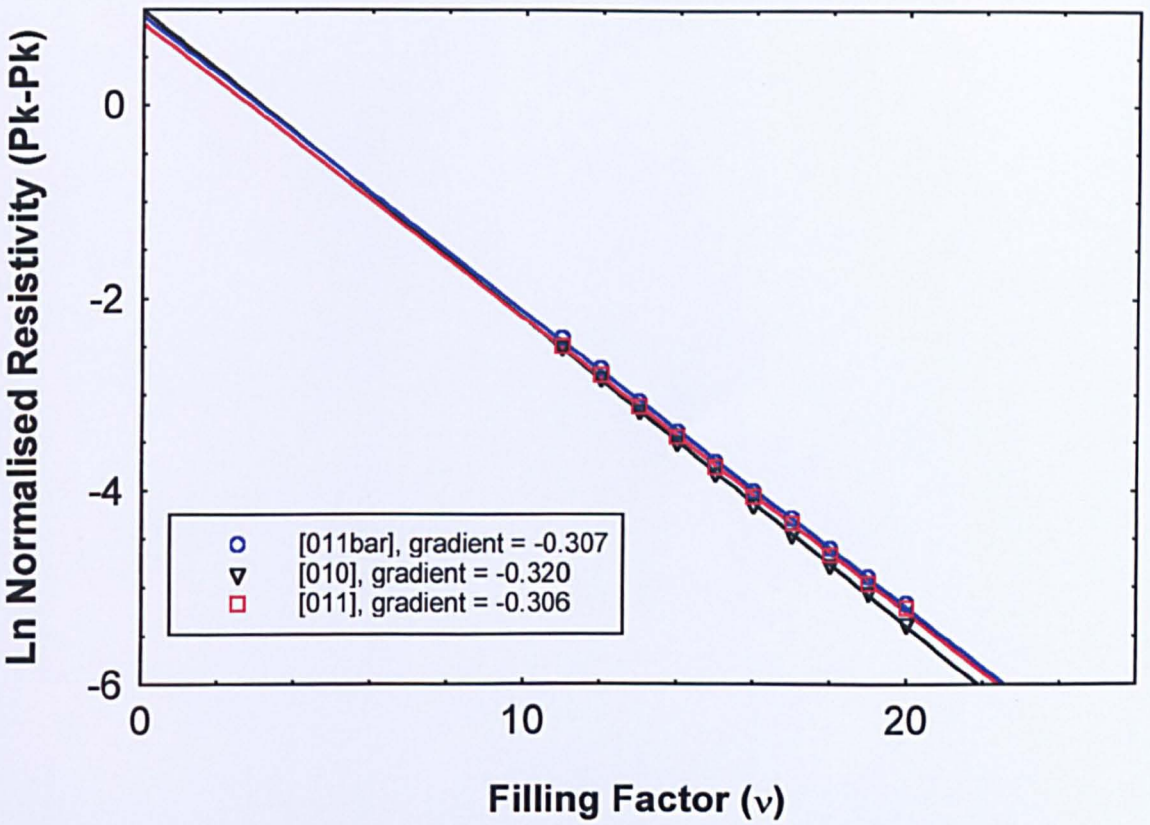


Figure 5.8.9 The Dingle plot corresponding to the curves in fig. 5.8.7.

mobility was confirmed as $(0.52\pm0.01)\times10^4 cm^2V^{-1}s^{-1}$, therefore its μ_{tr}/μ_q ratios are 6.7 and 5.2 in the directions $[01\bar{1}]$ and [010] respectively.

The variation of quantum mobility at carrier densities greater than that of the peak of sample A1236 was also studied.‡ A large range of gate bias was applied for good resolution and the corresponding set of SdH oscillations is shown in fig. 5.8.10. As with the transport mobility over an extended gate voltage range, the zero-field resistance reverses in value, but in this case there is a minimum. Fig. 5.8.11 shows the corresponding Dingle plot, where it is apparent that there is a monotonic *increase* in quantum mobility (decreasing gradient magnitude) as gate

‡ Data courtesy of Dr A.R. Long.

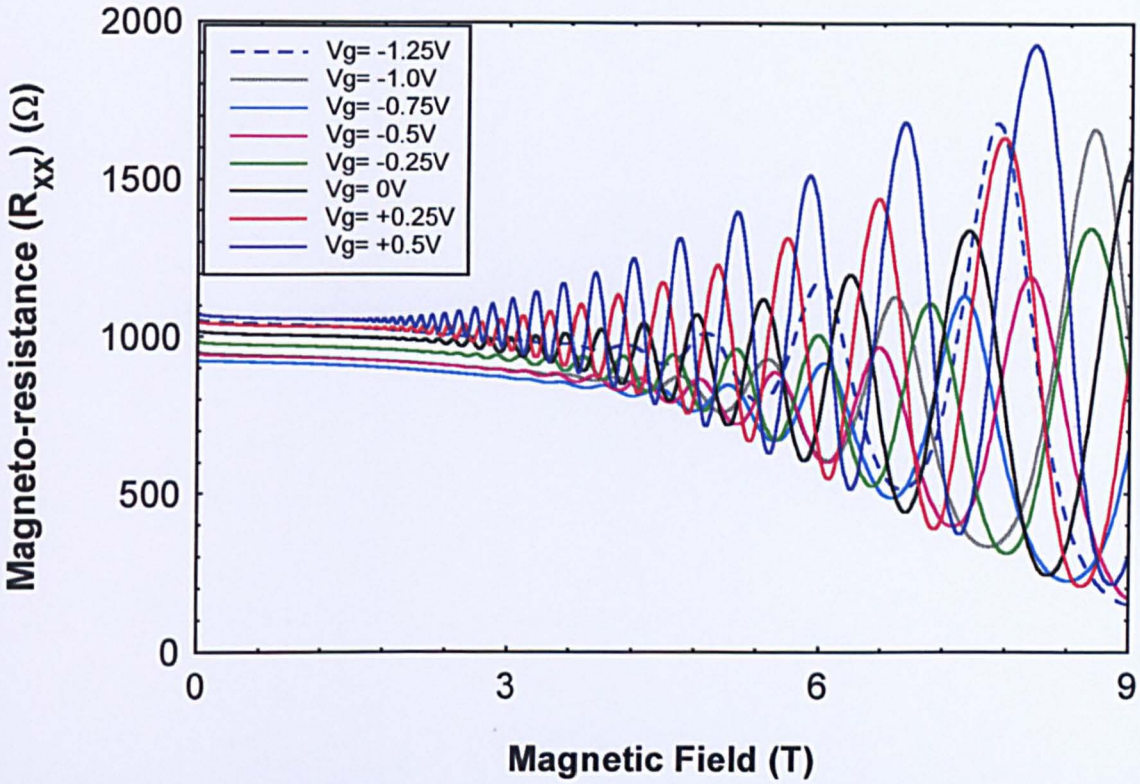


Figure 5.8.10 The Shubnikov-de Haas curves (1.6K) for sample A1236 (2.5nm spacer, 20% indium) over a range of gate voltages, covering carrier densities above the transport mobility peak, see fig. 5.3.1 or 5.5.1. Note the reversal of the zero-field resistance as the gate voltage is increased.

voltage is increased. This is contrary to the behaviour of transport mobility, which decreases, and zero-field resistance, which troughs over the same gate voltage range. This behaviour would be expected when there is a transition to a dominant isotropic scattering. This relative behaviour of the two measures of mobility was replicated in all the other 2.5nm samples.

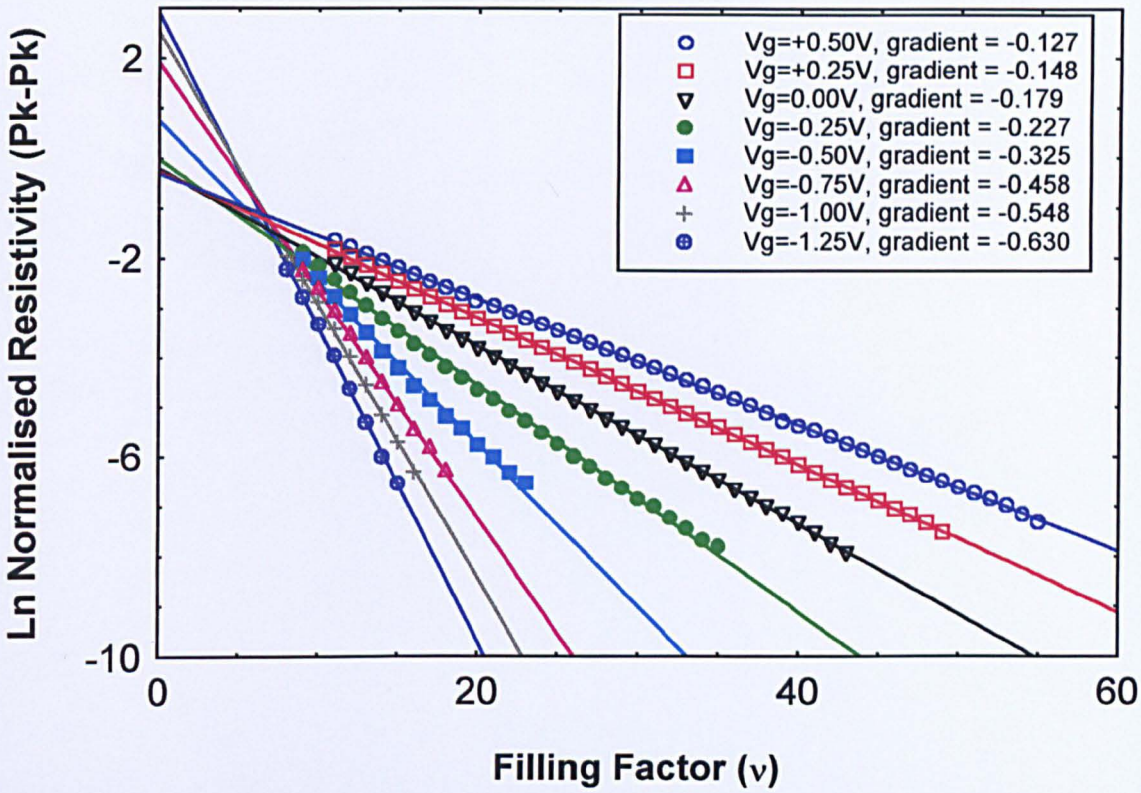


Figure 5.8.11 The Dingle plot corresponding to the curves in fig. 5.8.10. Note the continuous fall in the gradient magnitude as gate voltage increases, hence there is a continuous rise in quantum mobility, see equation 5.8.3.

B The Localised Regimes

5.9 The Localised Regimes

5.9.1 Introduction

The prospect of significant electron densities, somehow confined within the channel and hence unavailable for conduction, was inferred from measurements in the metallic regime, see fig. 5.3.1. Extrapolation of the curves to the zero mobility cut-off suggested that a sizeable carrier density (about $4x10^{11}cm^{-2}$) might be trapped in potential fluctuations at the conduction band-edge. Extrapolation was necessary as 4-terminal measurements fail at high channel depletion. In this configuration the gate covers all of the Hallbar voltage probes and the current terminals are elsewhere at the extreme ends of the Hallbar. Therefore, one or both will be subjected eventually to local depletion of electrons. It was decided that 2-terminal conductance measurements, where the current is also supplied between the same two terminals that measure voltage, should clarify the transport in this regime. To simplify the geometry, the terminals at the extreme ends of the Hallbar were used. Additionally, electrical measurements would complement the microscopic evidence[3] revealing more about the profile of the potential fluctuations.

As well as exposing the strongly localised regime (section 2.5.3) to greater scrutiny, the 2-terminal measurements also revealed that weak localisation (section 2.5.2) may be present. (4-terminal measurements at

variable temperatures in the metallic regime also confirmed weak localisation characteristics at similar carrier density.) Weak localisation data acquired 2-terminally was obtained at slightly higher carrier densities than those where strong localisation was observed. As the channel is depleted, the Fermi level sinks lower in energy and it would be expected that the wavefunction of the electrons would interact more strongly with the potential fluctuations. If this energy is just above that of the fluctuations, then the supposition is that the wavefunction will be strongly influenced by them. On further depletion, the Fermi level sinks within the fluctuations thus requiring the electrons to be thermally activated or to tunnel if they are going to contribute to conduction.

2-terminal measurements were thus made over negative gate voltages, such that both the strong and weak localisation regimes were encompassed. To test over a large range of activation energy, measurements were made over a large temperature range, between about 1.7K and 100K. Additionally, 4-terminal magneto-resistance data at various temperatures and carrier densities (gate voltages) was acquired to complement both the variable temperature Hall data and the conductivity data in the weak localisation regime. The 2-terminal measurements will be discussed first.

5.9.2 General 2 -Terminal Observations

A typical set of plots relating the various parameters of conductivity, gate voltage (and hence carrier density) and temperature for sample A1235 is shown in figure 5.9.1. The data set from which all other pertinent relations were extracted is shown towards the RHS of the plot. This shows how the square conductivity diminishes with gate voltage (hence carrier

density) at various temperatures from about 1.7K to 98K. In this form, these curves lose their potential for yielding useful information at gate voltages less than about -3.0V. Beyond this value, the strong localised regime is approached and entered, where the square conductivity

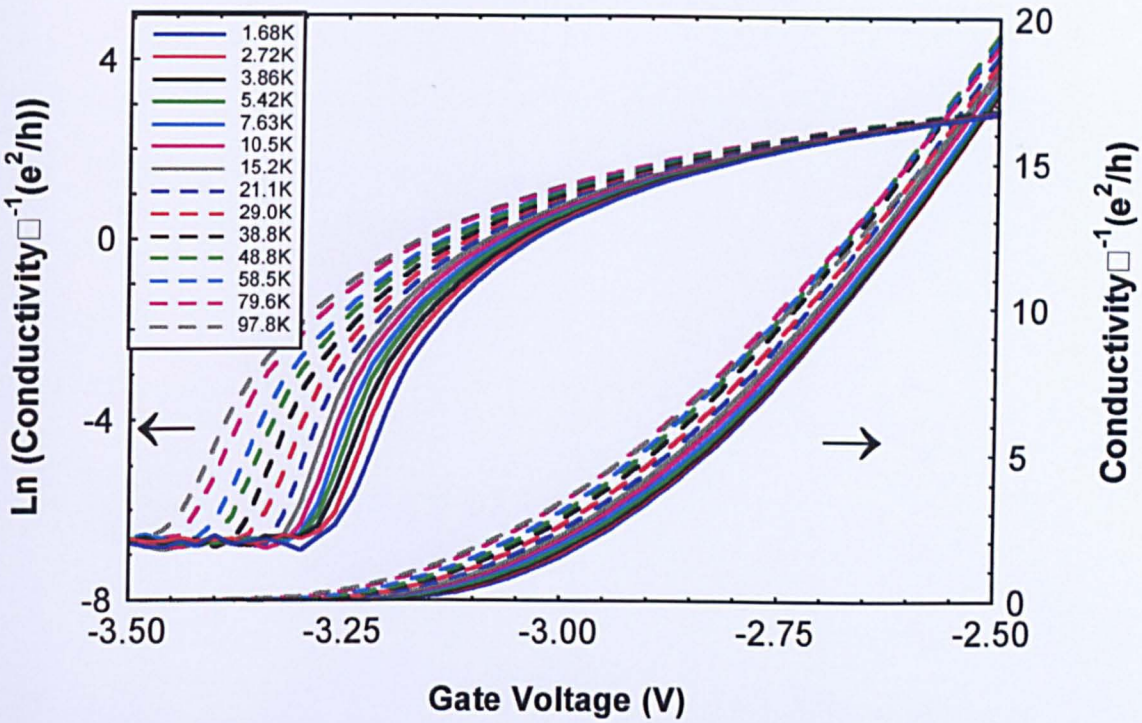


Figure 5.9.1 The plots of the square conductivity (RHS) and its natural logarithm (LHS) versus gate voltage for sample A1235 (20% indium, 5nm spacer). The data was acquired 2-terminally between temperatures of 1.7K and 98K. At gate voltages less than about –3.0V, the strong localising regime is entered. Above this value the metallic regime is accessed.

becomes less than about e^2/h. To clarify this region, the natural logarithm of conductivity is calculated and this is plotted on the left side of the graph. Note that in this form the curves 'fan' out and actually have quite large rates of change of conductivity. Although the actual magnitudes are small, these rates were totally obscured in the original unprocessed data set. A "saturated" (minimum) conductivity or current is also apparent and represents the minimum measurable current sustainable in a given layer under the appropriate conditions. DC experiments gave a similar but

slightly smaller residual current, which was also noisier. However, the relevant data was in agreement.

The conductivity data set is more useful at gate voltages greater than -3.0V (more positive values), where a changeover to the metallic regime would be expected. Note that the apparent divergence of the curves with temperature ceases about here and the lines become generally parallel to each other. This is indicative of changes in conductivity being independent of the conductivity itself - equation 2.5.14 shows that $\Delta\sigma$ is approximately independent of σ as the temperature changes - an experimental sign of the presence of weak localisation. In other words, $d\sigma/dV_g$ is independent of σ as a function of temperature.

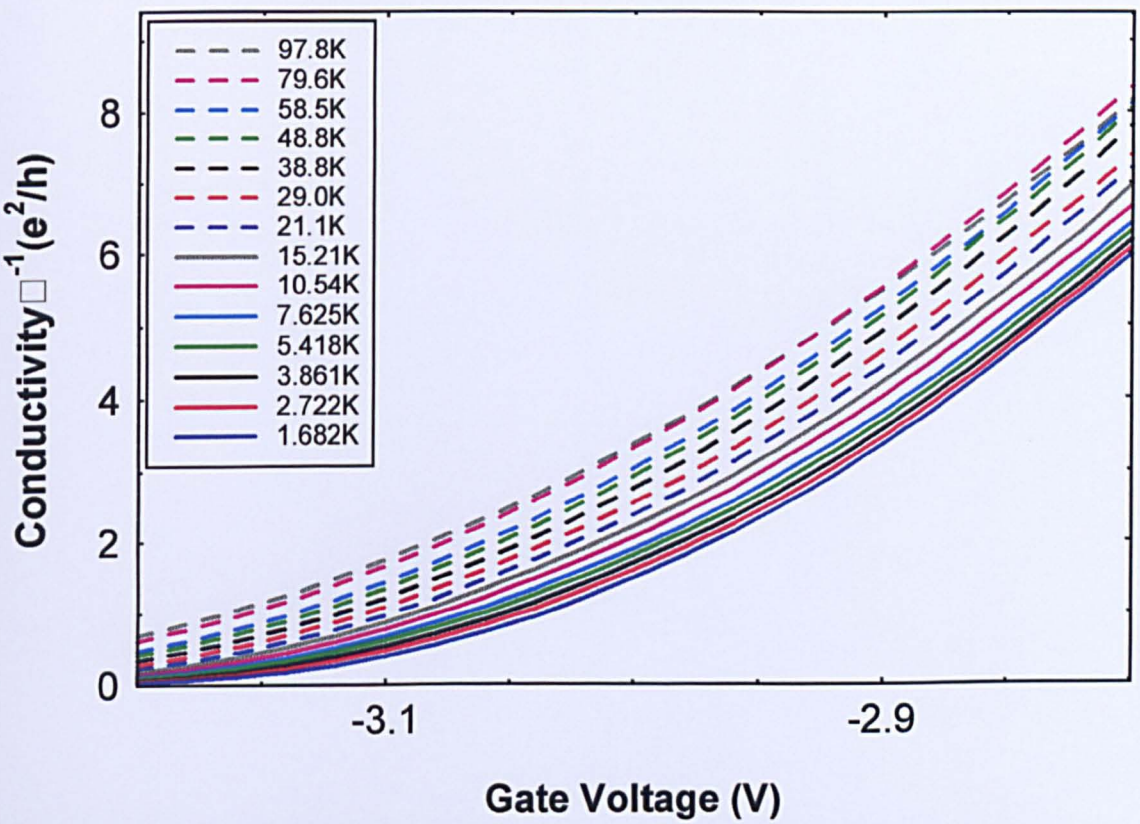


Figure 5.9.2 A magnified view of the square conductivity data from the RHS of fig. 5.9.1 showing the transition between the metallic and strongly localised regimes around –3.0V. Note how the curve representing the highest temperature drops below that of the 79.6K example at the highest gate voltage shown.

As the carrier density increases, the high temperature conductivity curves begin to overlap those for lower temperature. This is expected when weak localisation is destroyed, as the higher temperature will lead to a much-reduced phase breaking length and hence limit the conductivity. This effect will be more significant for higher conductivity as the elastic mean free path will be larger. At these temperatures therefore, coherent back scattering becomes difficult to maintain and weak localisation is quickly destroyed. As carrier density increases further, more curves representing the lower temperatures also begin to converge with the lowest temperature data. Over the range shown (fig. 5.9.1) however, weak localisation effects seem to be maintained provided a temperature of 20K is not exceeded. A magnified view around the transition between the strongly localised and the metallic regimes is shown in fig. 5.9.2.

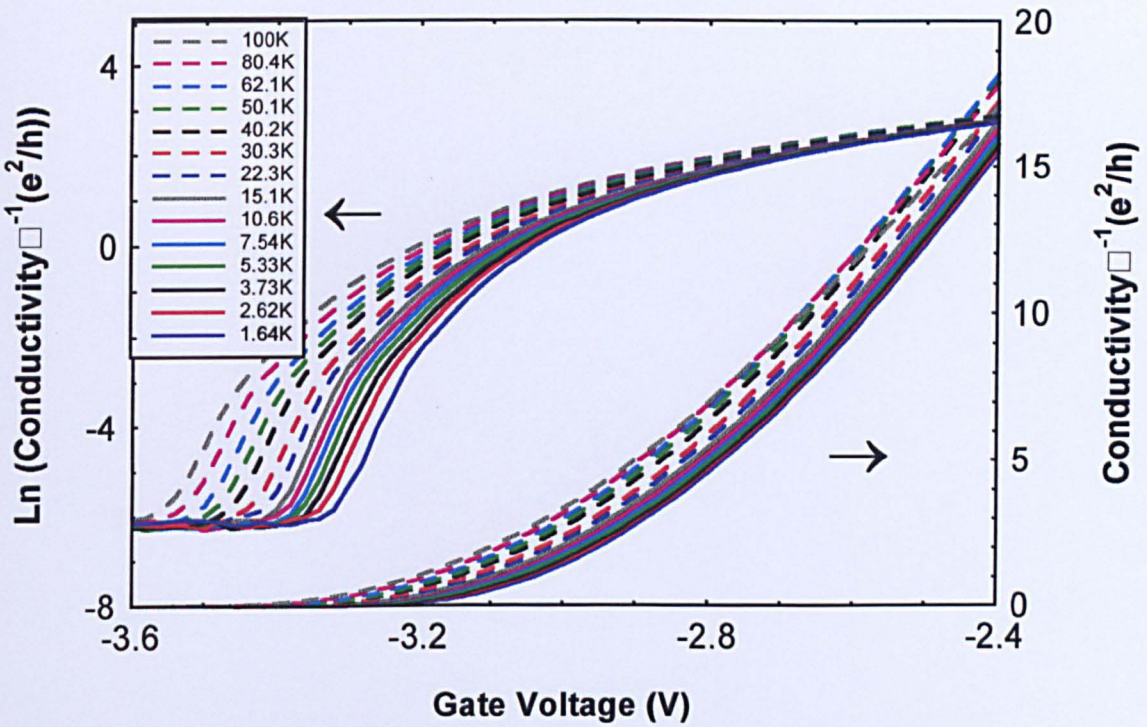


Figure 5.9.3 The same parameters plotted for sample A1236 (20% indium, 2.5nm spacer) as plotted for sample A1235 in fig. 5.9.1.

Another illustration of this is shown in fig. 5.9.3 for A1236, where at the maximum gate voltage shown similar behaviour (the data for the highest

temperatures dropping below that for lower temperature curves) is also apparent. The data for a sample (A1194) with a shallow channel is also shown in fig. 5.9.4, although the conductivity range covered is smaller.

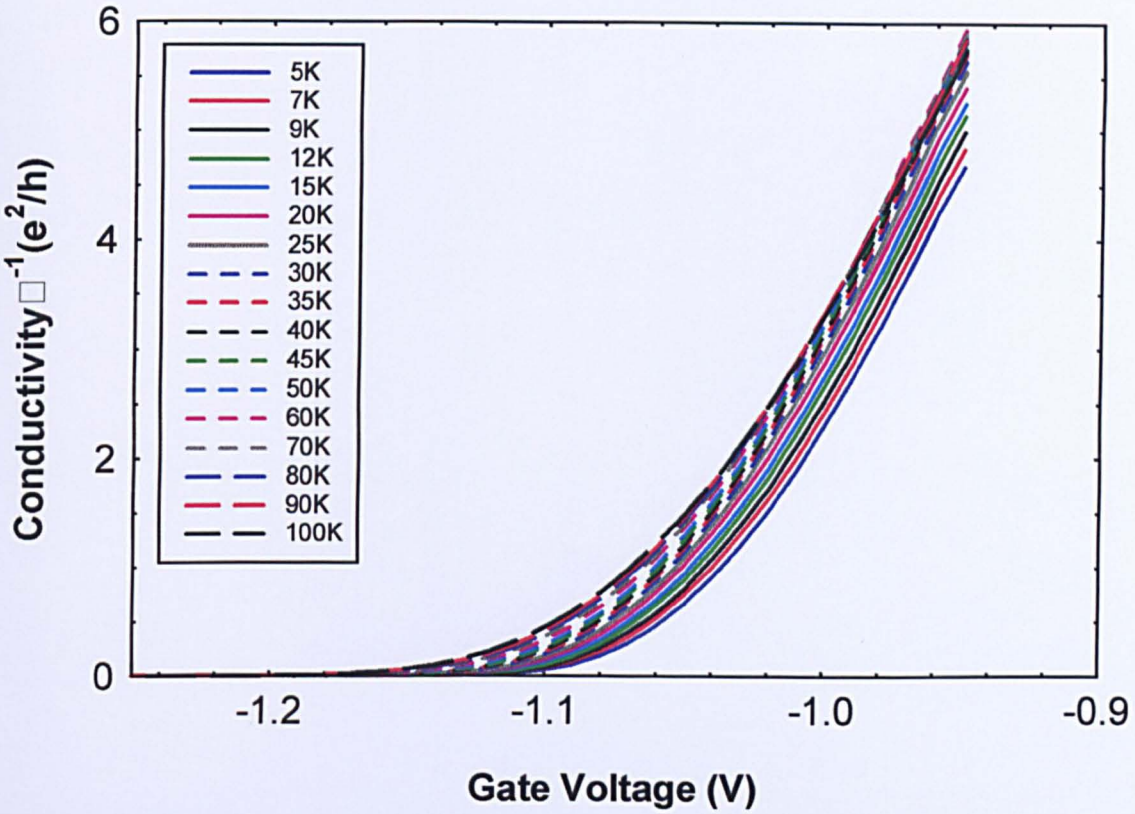


Figure 5.9.4 The gate voltage dependent square conductivity for sample A1194 (30% indium, 10nm spacer; shallow channel – less negative gate voltage for depletion necessary).

From the conductivity versus gate voltage plots in figs. 5.9.1, 2 and 3, it is apparent that the curves representing the top temperature – around 100K in each case – begin crossing the lower temperature curves at different conductivities. The thinnest spacer specimen (A1236) crosses at the highest value – about 7.5 e^2/h, whereas the thickest (A1194) crosses at about 2 e^2/h. A1235 (5nm spacer) crosses at about a value of 5 e^2/h.

In order to resolve the nature of the transport in the strong localisation regime, the conductivity data shown in fig. 5.9.1 was transposed from a gate voltage dependent conductivity over a range of

temperature to a temperature dependent conductivity over a range of gate voltage (or carrier density). This so-called Arrhenius plot, where the log of the square conductivity is plotted versus the reciprocal of the temperature, is shown in fig. 5.9.5. The general behaviour at most of the gate voltages or carrier densities is that there is a gradual rise in the rate of increase of conductivity with increasing temperature. Saturation develops in this rate of rise towards the highest temperatures. Equation 2.5.15 shows that the gradient of these curves is related to the activation energy (E_A).

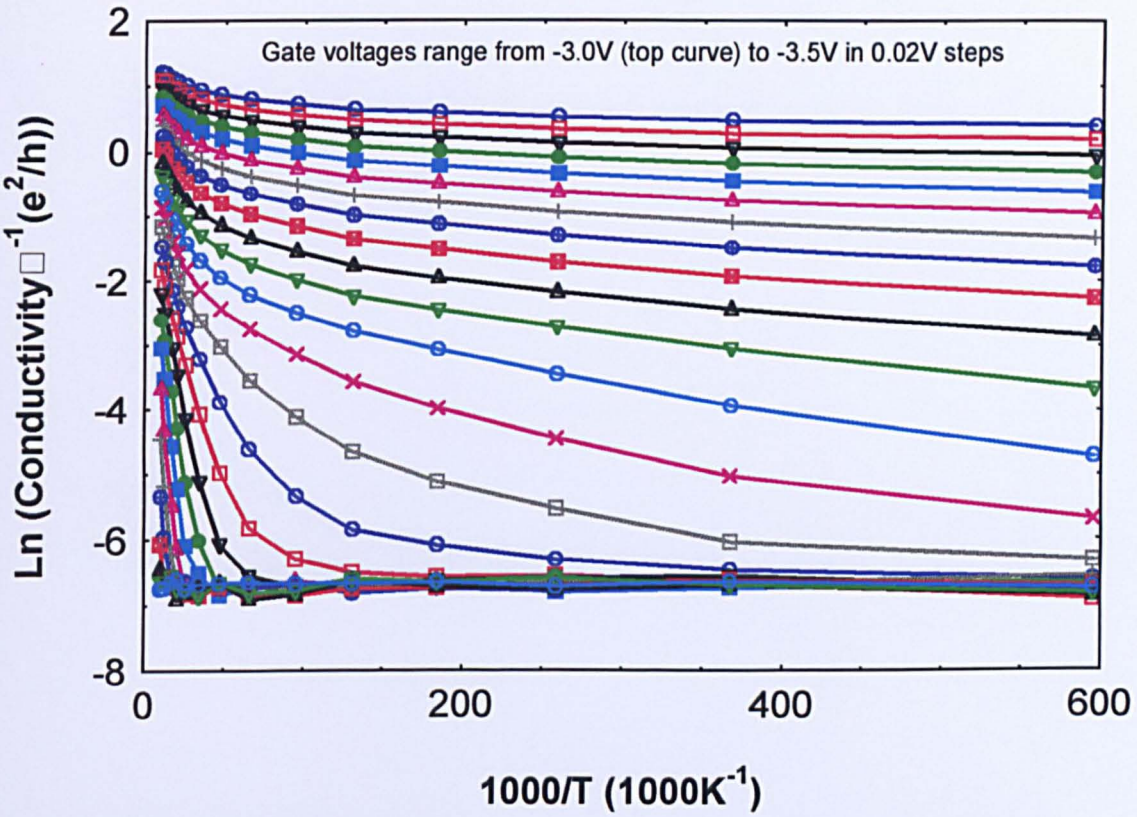


Figure 5.9.5 The Arrhenius plot of the transposed data from fig. 5.9.1 for sample A1235. Note that as temperature rises the gradients of most curves increase until saturation is apparent. These gradients are a measure of the temperature-derived thermal or activation energy E_A, calculated from equation 2.5.15.

This rate of increase is strongly gate voltage (carrier density) dependent and can be explained by considering smaller intervals of gate voltage or

carrier density starting with the highest. To accomplish this, it may be helpful to refer to both of figs. 5.9.5 and 5.9.6 – the latter showing a close-up around the higher temperatures with lines extrapolating to the ordinate axis. The curves representing higher carrier density are those which extrapolate at infinite temperature to a square conductivity greater than e^2/h (or zero on this log axis). This grouping is representative of the crossover between the strong and weak localised regimes for the following reasons:

1) A square conductivity above about e^2/h is largely accepted as being metallic in nature.
2) The curves of this set are becoming more parallel and flatter with increasing gate voltage – this is particularly noticeable at lower temperature. This observation is of course related to the parallelism mentioned previously.
3) The curves do not extrapolate to a single conductivity therefore there is no single mobility-edge, as defined in section 2.5. Again, this is a mark of metallic behaviour, or at least behaviour that is not characteristic of strong confinement where a single mobility-edge would be expected. This latter feature is normally associated with the activation energy, required to remove electrons from regions of strong confinement. At higher energy, above any confinement, this is no longer required.

As the carrier density decreases, the behaviour of another range of curves in this figure can be considered. These are the examples which extrapolate to a conductivity less than e^2/h. They are characterised by much larger rates of increase in conductivity where the electrons are thermally activated out of confining potentials. This is demonstrated by the larger gradients evolving at higher temperature and lower carrier density. This was also observed in the Si MOSFETs.[4] There is also a tendency for

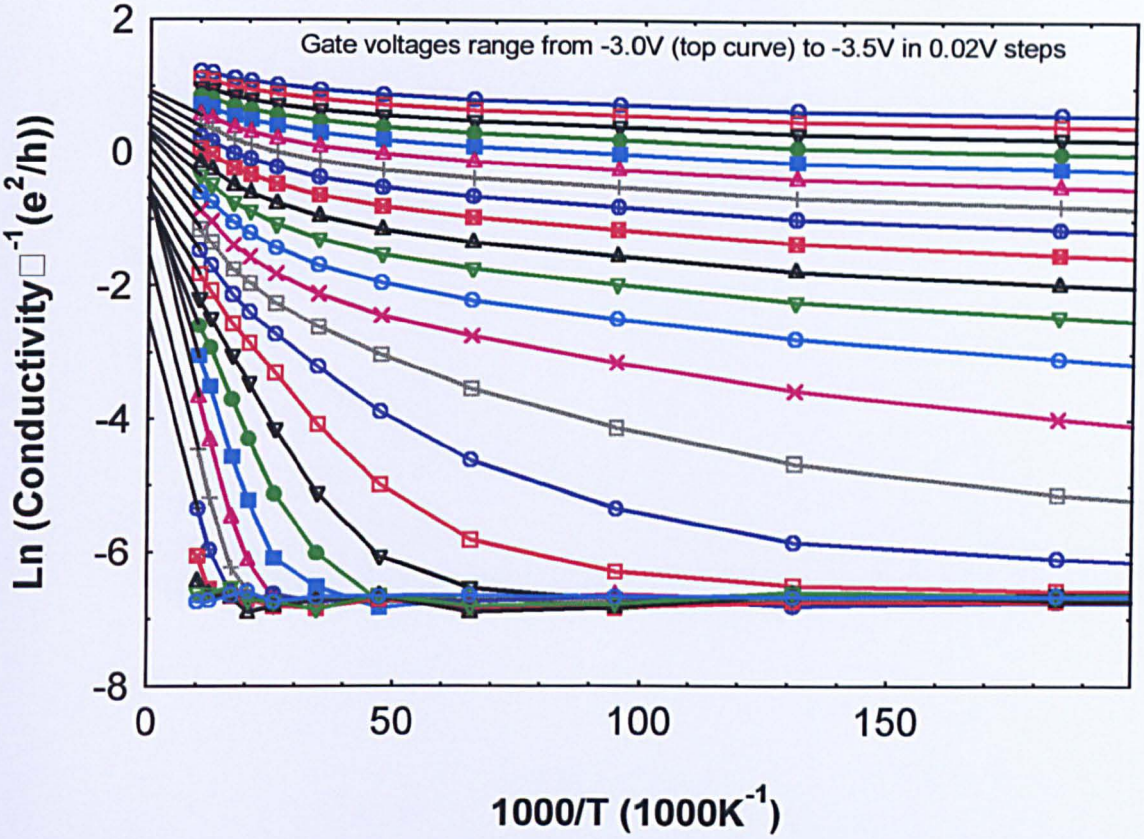


Figure 5.9.6 A closer view of the behaviour around the higher temperatures shown in fig. 5.9.5 for sample A1235. Lines are added to illustrate the extrapolation to the ordinate axis. The extrapolated values relate to σ_0 - the pre-factor in equation 2.5.15.

this set of curves to converge rather than remain parallel, although they do not necessarily converge to a single point. In most samples tested, there was convergence of some of the curves in this set but not all. In the example shown, six of the curves converge approximately to a single extrapolated conductivity (σ_0) ($(0.55\pm0.05)e^2/h$) – the prefactor in equation 2.5.15. The implication is that in this type of layer system no single mobility-edge can be constructed. This behaviour diverges from the so-called ideal as defined in section 2.5. Instead, the behaviour can be described as broadly non-ideal. Departure from the single-valued extrapolation tended to occur at higher conductivity rather than lower, although there was some sample and contact variability in this respect. More data at higher temperatures would have clarified the trend at high

depletion, but was not acquired, as the occupation of the *DX* centres would be changed. As stated earlier, values for the activation energy can be calculated from equation 2.5.15. As an example, these along with the corresponding extrapolated square conductivity values (σ_0) are shown for sample A1190 in fig. 5.9.7.

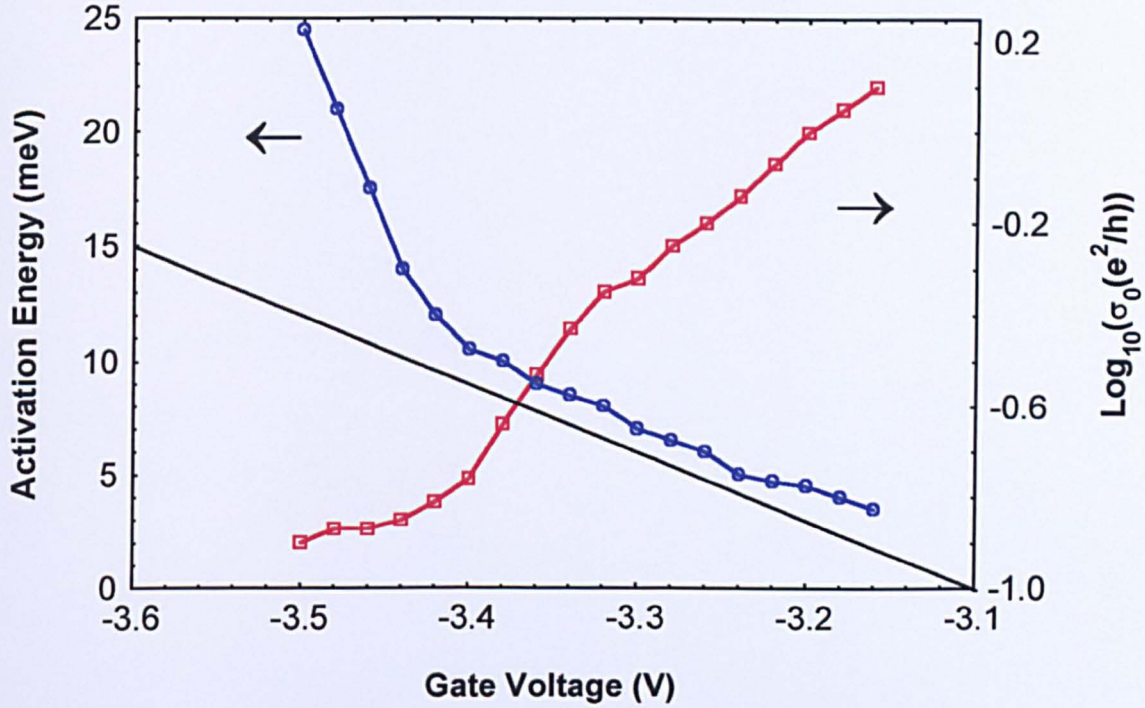


Figure 5.9.7 Plots showing the variability with depletion of the extrapolated conductivity and the corresponding activation energy for sample A1190 (30% indium, 2.5nm spacer). The straight line shown represents the rate of change of potential energy calculated using the full metallic 2D density of states. It has a gradient of -0.03e and represents a limiting gradient for the rate of change of E_A as the density rises towards the metallic regime.

Note that at high depletion, where the activation energy rises relatively steeply, there is a region where the extrapolated conductivity is fairly constant (about $0.5e^2/h$) and is comparable in magnitude with the corresponding value for Si MOSFETs ($0.14e^2/h$).[4] However, as above, there is also some sample and contact variation. At the highest carrier density depletion (lowest Fermi level), it becomes increasingly difficult for

the available thermal energy at about 100K to excite the remaining electrons out, thus allowing them to contribute to normal conductivity above the leakage current. In these high temperature conditions, electron transport is characterised by a low-level saturated 'hopping'-like behaviour. At lower temperature, there is weak temperature dependence with the effective activation energy decreasing with temperature. Note that the gradient of this blue curve tends approximately towards that shown (black line) for the rate of change of the potential energy calculated using the full metallic 2D density of states. This appears to limit the rate of change of the activation energy as the system enters the metallic regime.

At even lower temperatures than those shown (down to 1.6K) there is a continuation of the above trend with even lower - almost flat – gradients existing. There is very little thermal activation with tunnelling generally dominating the conduction, but the current magnitude is higher than that found in the MOSFETs. A partial reason for the higher currents lies in the larger localisation radius in InGaAs when compared with Si. The much lower effective mass of an electron within InGaAs confers it with a larger localisation radius. To determine any appropriate conductance mechanism, such as 2D 'variable range hopping', which follows a $T^{1/3}$ law, the data from the Arrhenius plot was transformed appropriately. The resulting data proved inconclusive as to which mechanism was responsible for transport even down to 1.5K. The very weak temperature dependence meant that a larger temperature range was required – possibly even below 1K, which was unavailable during the period of the project.

If the potential fluctuations are of the order of 100nm in size, then at the onset of strong localisation, there should be a few tens of electrons within this dimension. This number would eventually condense into

individual 'pools' confined by the fluctuations and then either hopping or tunnelling would commence between them. These pools are expected to be quantitatively much larger but qualitatively similar to those proposed[8] to arise in unstrained 2DEGs due to the influence of remote ionised donors. As in the hopping theory for cermets,[9] the Coulomb charging energy within the pools must be included in any analysis, therefore activation energies at low temperatures may be dominated by this self-charging. Typical charging energies of about 1meV have been measured in fabricated surface-gated quantum dots,[10] which confine electrons within a similar dimension. This value compares reasonably with the activation energies measured here at low temperatures.

5.9.3 Further Tests of Weak Localisation

5.9.3.1 Introduction

In section 5.3, it was hypothesised that a sizeable carrier density was trapped in potential fluctuations within the channel, and in section 5.5, it was shown that this extrapolated density varied with temperature. It is likely therefore, that behaviour reminiscent of weak localisation would precede this when the Fermi energy is slightly higher. This likelihood was strengthened when the observed reduction in conductivity with decreasing temperature was of the appropriate magnitude. Further evidence was presented in section 5.9.2, where changes in conductivity with temperature appeared independent of the conductivity itself - as suggested by equation

2.5.14. In response to this indirect evidence, more pertinent experiments were carried out to test these observations.

5.9.3.2 Experimental Tests for Weak Localisation

The first test carried out was to determine the relationship between conductivity and temperature using equation 2.5.14, which linearly relates

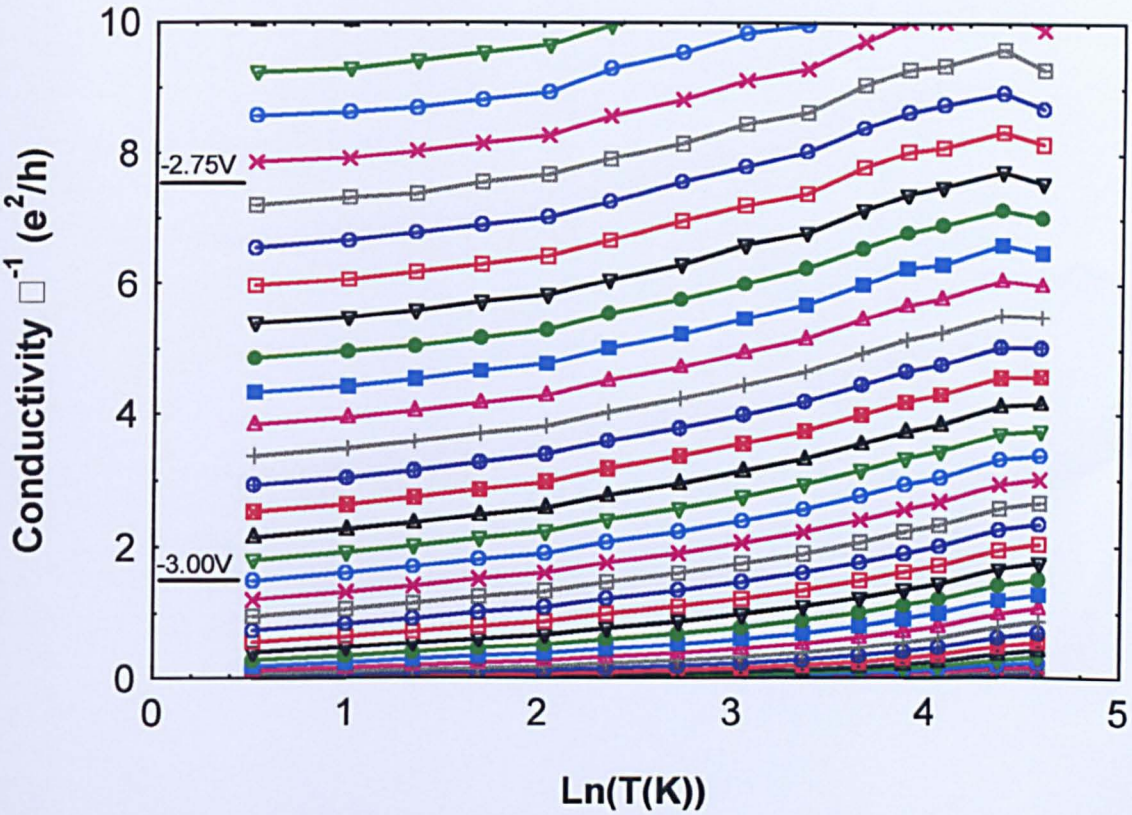


Figure 5.9.8 Plot of conductivity (σ) versus the natural logarithm of the temperature (lnT) for sample A1235. Each line represents a different gate voltage or carrier density - the values chosen to access the region where weak localisation was expected. The chosen gate voltages for analysis are marked, although –2.50V is off the top. No linear dependence was found over the complete temperature range, but reasonable linearity exists up to about 10K.

changes of the conductivity due to weak localisation with the natural log of the temperature. Two examples of this are shown in figs. 5.9.8 and 5.9.9 for samples A1235 and A1236 respectively, where each curve represents a

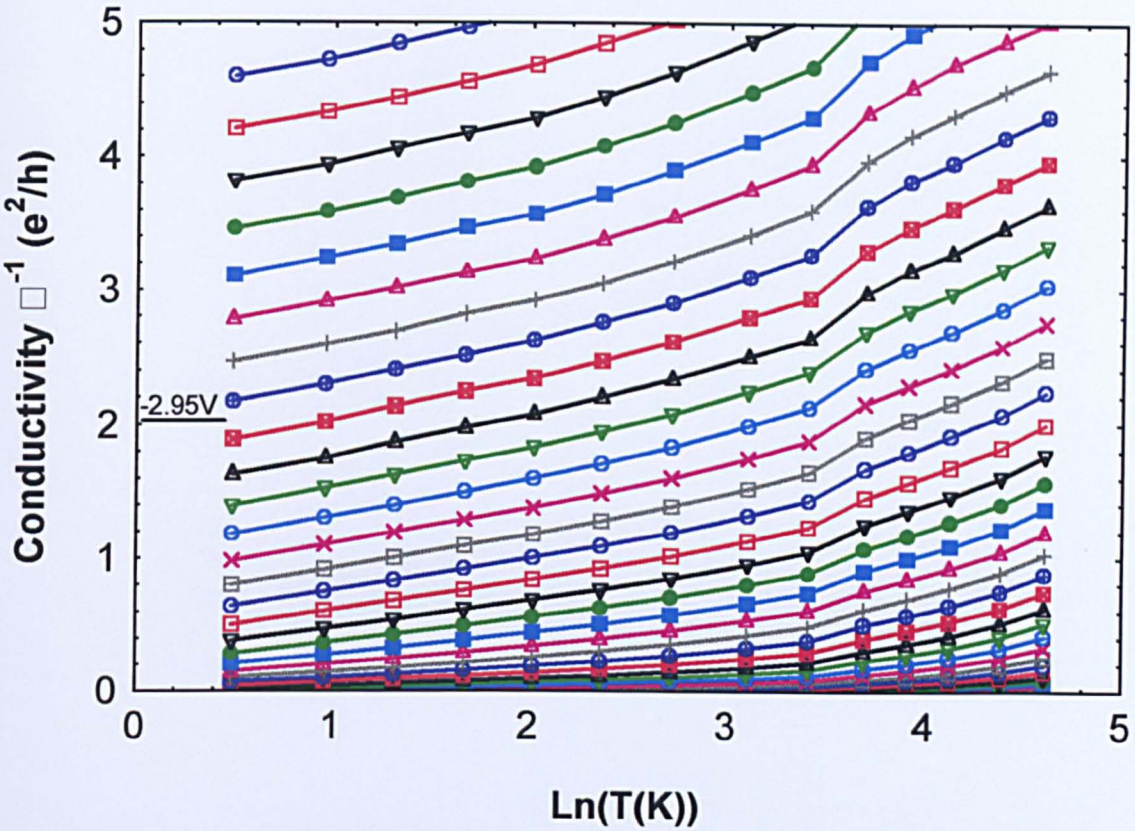


Figure 5.9.9 Plot of conductivity (σ) versus the natural logarithm of the temperature (lnT) for sample A1236. Each line represents a different gate voltage or carrier density - the values chosen to access the region where weak localisation was expected. One of the gate voltages used for analysis is marked, but the other (-2.70V) is off the top of the graph. No linear dependence was found over the complete temperature range, but reasonable linearity exists up to about 35K – a greater range than that for sample A1235.

different gate voltage or carrier density. No convincingly linear relationship was found over the complete temperature range in any of the samples tested. There was better linearity at low temperatures. However some samples, for example A1236 (fig. 5.9.9), displayed better linearity

than others (A1235 – fig. 5.9.8). (If instead of σ, $ln\sigma$ is plotted versus lnT, then it is found that the curvature (2nd derivative) of the lines changes

Sample	Vg(V)	T(K)	μ	n	B_{Tr} (T)	τ_ϕ/τ	σ(T)	DT	Δσ(B)	DB
A1235	-2.50	1.72	0.828	0.539	0.0327	16	55.0	1.5	-2.77	1.17
		9.3	0.858	0.545	0.0301	5	56.5		-1.6	
	-2.75	1.72	0.448	0.400	0.151	28	23.8	1.8	-3.33	1.54
		9.3	0.486	0.409	0.125	6	25.6		-1.79	
	-3.00	9.3	0.165	0.278	1.60	20	6.25	2.35	-3.00	1.90
		38	0.260	0.240	0.745	3	8.60		-1.10	
A1236	-2.70	1.67	0.428	0.620	0.106	36	21.5	2.2	-3.58	1.88
		12	0.457	0.623	0.0929	5.5	23.7		-1.70	
	-2.95	1.67	0.199	0.499	0.612	60	6.44	1.94	-4.09	1.61
		12	0.236	0.498	0.436	12	8.38		-2.48	

Table 5.9.1 A table of experimental and calculated theoretical parameters relevant to a study of weak localisation affected by temperature (T) and magnetic field (B). The samples used were A1235 and A1236. **KEY:**-Vg - gate voltage; T – temperature; μ- transport mobility ($x10^4 cm^{-2}V^{-1}s^{-1}$); n- carrier density ($x10^{12}cm^{-2}$); B_{Tr} – characteristic transport magnetic field; τ_ϕ/τ - ratio of phase to transport relaxation time; Δσ - conductivity correction for temperature (T) and magnetic field (B) in units ($e^2/\pi h$). DT refers to the difference in conductivity between the two temperatures at a given gate voltage – the numbers taken from the column on the immediate left. DB represents the same procedure but applied to the column on its immediate left. The temperatures shown are those where a measurement could be made from both the temperature data- figs. 5.9.8 and 9 and the magnetic field data – figs. 5.9.11-13. σ(T) is the value of the conductivity at a given temperature and gate voltage taken directly from figs. 5.9.8 and 9, and Δσ(B) is the weak localisation correction at a given temperature based on fitting the experimental curves at a given temperature and gate voltage to the theoretical curves using the ratio τ_ϕ/τ as a fitting parameter (see equation 5.9.2) and B_{Tr} calculated using equation 5.9.1, with n and μ values taken from other experimental data. With the τ_ϕ/τ ratio determined, Δσ(B) can be calculated using equation 2.5.12.

sign at a gate voltage approximately commensurate with the changeover between the strong and weak localised regions) However, the conductivity (σ(T)) at a given temperature and gate voltage can be extracted and the values are given in table 5.9.1, (column headed σ(T)), where they have been multiplied by the factor π for direct comparison with

weak localisation changes induced by a magnetic field (see later). The column to the immediate right (DT) contains the difference in the conductivity between two successive temperatures – the two values to the left – derived from the temperature data.

As stated in section 2.5, probing the systems with a magnetic field can also produce information regarding weak localisation. A magnetic field will destroy weak localisation through the breakdown of time-reversal invariance and there are various magnetic fields characteristic of the particular mechanism responsible for its destruction, see equation 2.5.3. In this work a perpendicular field was applied and the magneto-resistance noted at various low-valued carrier densities and temperatures. Equations 2.5.4 and 2.5.10 or acceptable approximations, can provide a theoretical comparison with respect to determining the contribution of weak localisation in the layers of this work. The transport characteristic magnetic field (B_{Tr}) is determined from the available experimental parameters (n and μ) (see also table 5.9.1):

$$B_{tr} = \frac{\hbar}{4eD\tau} = \frac{\hbar}{4Dm^{*}\mu} = \frac{\hbar}{2el^{2}} = \frac{e}{2hn\mu^{2}} \qquad (5.9.1)$$

where D is the diffusion coefficient which is equal to $v_F^2\tau/2$ or $l^2/2\tau$ (v_F being the Fermi velocity, τ the transport scattering time and l the mean free path); m^* is the effective electronic mass; n is the carrier density and μ is the mobility which is equal to $e\tau/m^*$. Calculated values for B_{Tr} are shown in table 5.9.1, using equation 5.9.1 (the right-hand term) for two different samples – A1235 and A1236.

By ignoring the spin-orbit characteristic magnetic field (B_{so}) and using an appropriate approximation, the theoretical equation 2.5.10 for the weak localisation correction (Δσ) can be simplified to:

$$\Delta\sigma = \Psi\left(\frac{1}{2}+H\right) - \Psi\left(\frac{1}{2}+\frac{B_{Tr}}{B}\right) + \ln\left(\frac{\tau_\phi}{\tau}\right) \qquad (5.9.2)$$

where τ and τ_ϕ are the transport and phase relaxation times respectively and H is:

$$H = \frac{B_{Tr}}{B}\frac{\tau}{\tau_\phi} \qquad (5.9.3)$$

The ratio of τ_ϕ to τ can be set to fit, as best as possible, the experimental magneto-resistance or magneto-conductivity data. The normalised magneto-conductivity is plotted in fig. 5.9.10 for sample A1235 for two different temperatures – 1.72 and 9.3K at a gate voltage of –2.50V
Also plotted alongside each experimental curve is the theoretical 'best' fit to the data at low magnetic field using equation 5.9.2 and 5.9.3; a calculated value for B_{Tr} from equation 5.9.1 and an appropriate value for the ratio of τ_ϕ/τ. Note that, in both cases, the experimental data fits the theory quite well at very low fields, but soon diverges at higher fields with a positive magneto-conductance and hence violates the theory. At a gate voltage of –2.50V, the conductance is perhaps too high (55 $e^2/\pi h$ from the temperature study – fig. 5.9.8) for significant weak localisation to be detected. The low values of B_{Tr} at these two temperatures (about 0.03T) signify that weak localisation is quickly destroyed at low fields. With the ratio of τ_ϕ/τ determined from the theoretical fit, the equation 2.5.12 can now be used to calculate the value for the magnetic weak localisation correction ($\Delta\sigma(B)$), and this value is put in table 5.9.1 under the $\Delta\sigma(B)$ heading. In the column headed DB, are the **d**ifference values between weak localisation values found at successive temperatures using the **B** field technique. A comparison with the data in the DT column is now possible The mechanism behind the persistent positive magneto-

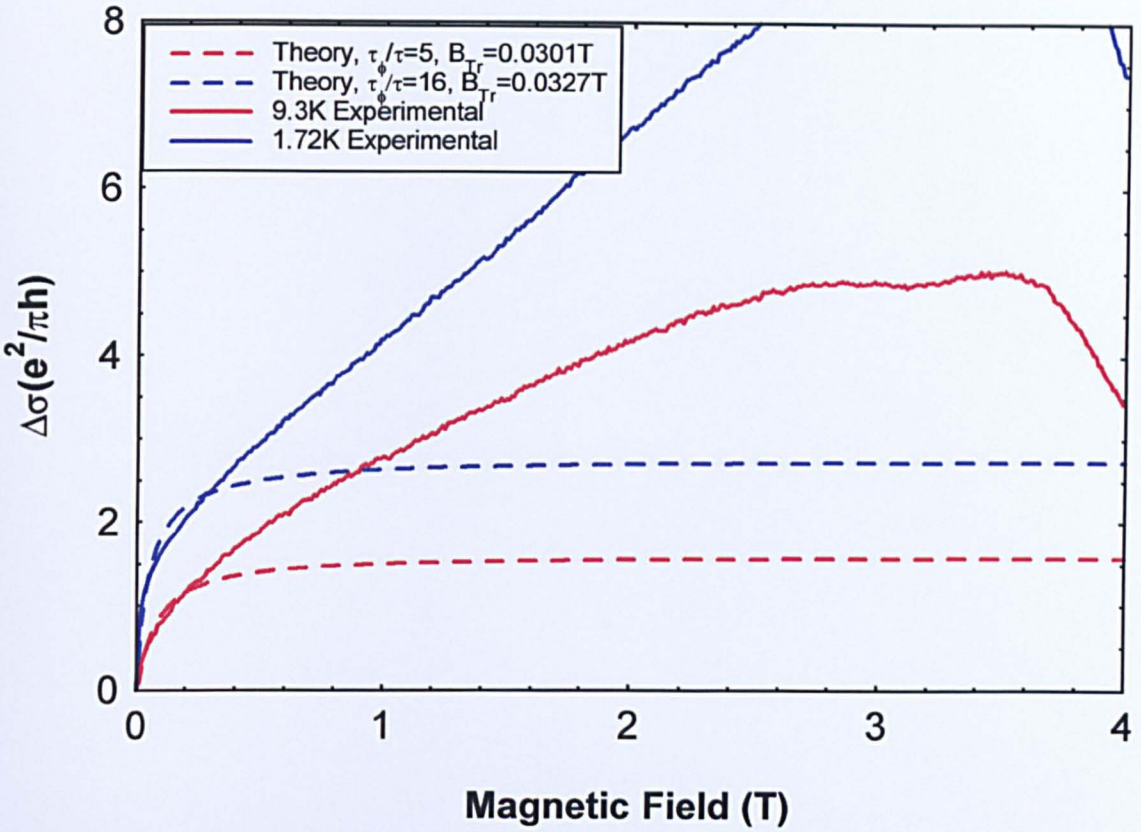


Figure 5.9.10 The comparison of the experimental and theoretical conductivity changes ($\Delta\sigma$) in response to a magnetic field at a gate voltage of –2.50 volts for sample A1235. The experimental curves are solid and the corresponding theoretical curves are dashed using the same colour. The appropriate characteristic fields are given in the legend.

conductance is unknown, but it was a common feature found in the data.

A similar plot for A1235, but with a gate voltage of –2.75V is shown for the same two temperatures in fig. 5.9.11. A similar plot for sample A1236 at one gate voltage (-2.70V) and two temperatures (1.67K and 12K) is shown in fig. 5.9.12. A study of the results in table 5.9.1, indicates that there is a reasonable correlation between the temperature

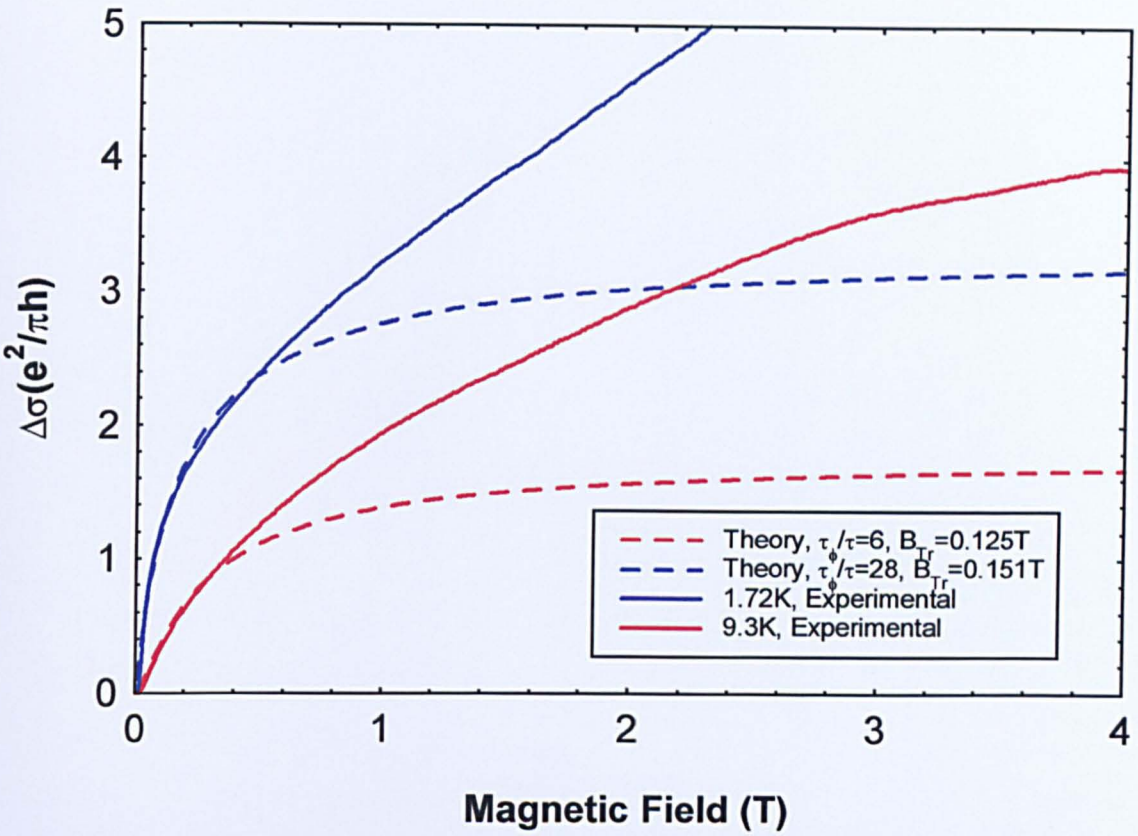


Figure 5.9.11 A similar plot as presented as fig. 5.9.10, but with a gate voltage of –2.75V. Note the better fit to higher fields indicating more significant weak localisation.

and magnetic field weak localisation corrections (DT) and (DB) respectively. At low temperatures generally, it is found that there is less than a factor of a third difference between the two changes in conductivity. At higher temperatures, (greater than about 15K) the difference is much greater - indeed, at the highest temperatures, it was impossible to fit the theoretical curves to experiment. Therefore, provided the conditions of temperature and gate voltage (carrier density) were favourable, then weak localisation is at least significant in these conditions, but other mechanisms exist at higher magnetic field.

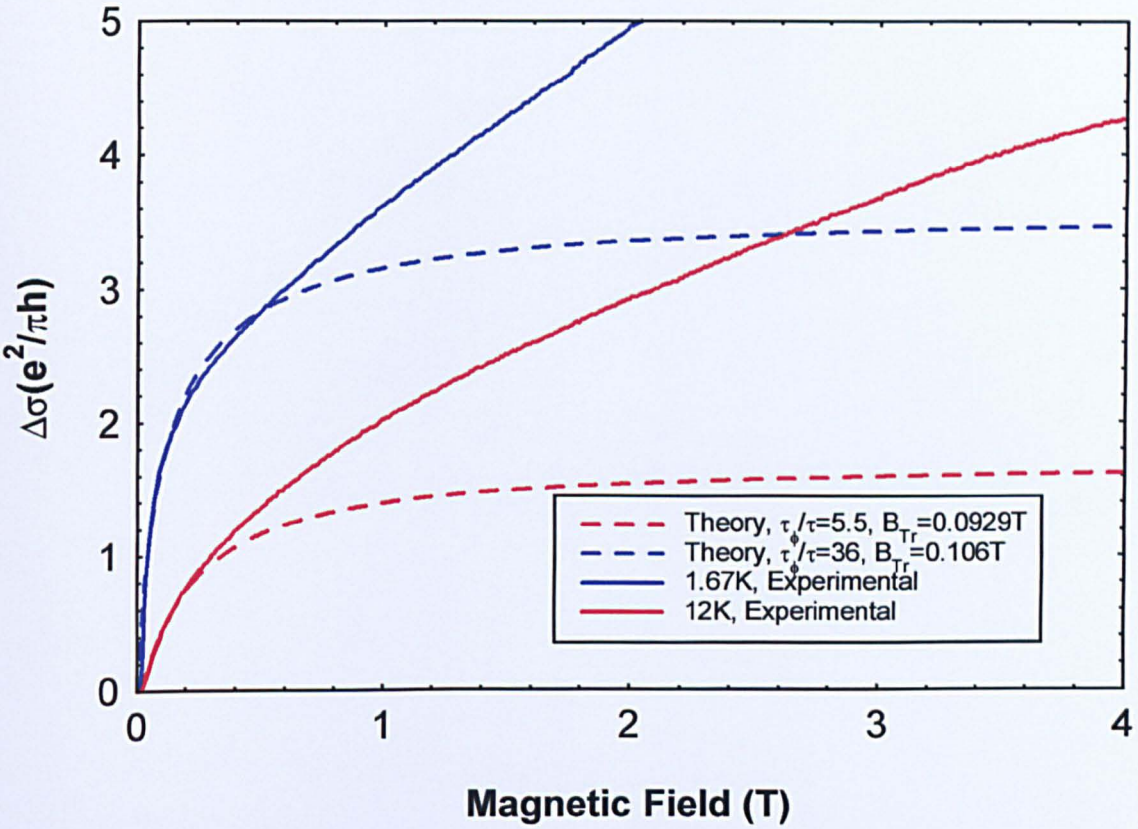


Figure 5.9.12 A similar plot as fig. 5.9.10, but for sample A1236 at a gate voltage of –2.70V. There is a similar degree of fit between experiment and theory as sample A1235 at a gate voltage of –2.75V, see fig. 5.9.11.

5.10 References

1 Mooney P.M., Semiconductior Science and Technology, **6**, B1, (1991).

2 Marsh J.H., Applied Physics Letters, **41**, 732, (1982).

3 Cullis A.G., Pidduck A.J., Emeny M.T., Journal of Crystal Growth **158**, 15, (1996).

4 Ando T., Fowler A.B., Stern F., Review of Modern Physics, **54**, 437, (1982).

5 Weisbuch C., Vinter V., Quantum Semiconductor Structures; Fundamentals and Applications, Academic Press, (1991).

6 Schweizer T., Kohler K., Rothemund W., Ganser P., Applied Physics Letters, **59**, 2736, (1991).

7 Beck M., Ilegems M., Proceedings – International Conference on InP and Related Materials, (1996).

8 Nixon J.A., Davies J.H., Physical Review B, **41**, 7929, (1990).

9 Neugebauer C.A., Webb M.B., Journal of Applied Physics, **33**. 74, (1962).

10 Borsosfoldi Z, Rahman M., Larkin I.A., Long A.R., Davies J.H., Weaver J.M.R., Holland M.C., Williamson J.G., Applied Physics Letters, **66**, 3666, (1995).

CHAPTER 6

CONCLUSIONS

MOTIVATION AND AIM

The extremely high mobility presently obtainable within modulation-doped GaAs/AlGaAs heterostructures at very low temperature has been a significant achievement. Many factors have contributed to this success including high quality MBE growth and well-tested fabrication procedures. However, if such structures are to form viable practical devices and circuits, the characteristics at room temperature must be acceptable. Present applications demand high conductivity with very high electron speeds or frequencies coupled with low noise. Such conditions can be satisfied using material with a high carrier density, reasonable mobility and a low effective electron mass.

Much research has been undertaken to optimise both the MBE growth and fabrication procedures to achieve these characteristics. The contribution from MBE lies in the growth of a ternary channel, which may be lattice mismatched to the layer below. Provided the layer thickness does not exceed a critical value, any strain can be

accommodated elastically. One of the consequences, is a better lattice match to the subsequent spacer layer, thus allowing some degree of band-gap engineering and the establishment of large conduction band-edge offsets at the channel spacer interface. The result is a higher 2DEG concentration, and provided the mobility is not seriously degraded, an acceptably high conductivity will result. The system used to achieve that in this work is a pseudomorphically strained InGaAs channel grown on a GaAs substrate with an AlGaAs spacer separating the channel from a δ-doped layer.

This work continues the process development for this system, with an emphasis on the physical characterisation of the scattering and mobility found in such material and to provide relevant experimental data for device modelling. By varying certain crucial parameters such as the indium content and MBE growth temperature along with the spacer width and doping concentration, a reasonable range of preparation conditions can be tested. Moreover, use of a range of characterisation techniques enables us to probe deeply into the details of the scattering processes.

RESULTS

Extensive transport measurements have been completed on deep and shallow-channelled InGaAs p-HEMTs of varying growth temperature, indium content, spacer thickness and doping density, with a view to a thorough characterisation, both in the metallic and the localised regimes.

Applicability of an Electrostatic Model (sect. 5.2)

Each gated Hallbar was tested for the applicability of an electrostatic model (sect. 2.1 and 5.2), which predicts carrier density, threshold gate voltage and other parameters. This model pins the Fermi energy on the deep-donor levels of the doping layer. Excellent agreement was obtained for carrier density of normally doped samples, which to a first order was independent of doping density. However, the other parameters were affected by the degree of overdoping and resulting electron occupation of the shallow states of the doping layer. In particular, linear depletion of the channel electron was severely compromised by overdoping. By changing the deep-donor energy in the model, the predicted carrier density of under-doped samples agreed with experiment.

Scattering and Mobility at Low Temperature After a Normal Cool (sect. 5.3, 5.6 and 5.8)

At low carrier densities in the metallic regime, the mobility is dominated in all samples by an intrinsic scattering factor arising in the channel. Remote ionised impurity (RII) scattering reduces the mobility slightly, particularly for the samples with thinner spacers (2.5nm). The intrinsic factor is ascribed to the long-range potentials arising from the indium concentration fluctuations. These spatial potential variations are perturbations of the conduction band-edge, and hence band-gap, caused by the spatial clustering of indium in the channel layer, or are associated with thickness fluctuations in the channel. The mobility is generally higher for samples with a lower indium content.

For the thickest-spacer (10nm) samples, scattering from this source continues to be dominant at higher carrier density, but for

thinner-spacer (2.5nm) samples another significant factor emerges. This is a strong additional spacer-dependent factor, which limits transport mobility at high densities with a characteristic mobility peak emerging for the thinnest-spacer samples. However, quantum mobility rises monotonically over the same carrier density range past the peak where transport mobility falls. This shows the scattering becomes much more isotropic in this range. The mobility peak is equally strong for a sample with a 20% indium content as it is for samples with a content of 30%. Although more study is required, direct resonant scattering of the tail of the wavefunction of the channel electrons by states within the donor layer is thought to be the cause of this behaviour.

Evidence of RII scattering exists at all carrier densities if the spacer width allows, but it is never thought to dominate the scattering characteristic of any sample at any density. Random alloy scattering has not been conclusively identified in the data. There is strong evidence of a large confined density (2-8x$10^{11}$$cm^{-2}$ does not contribute to metallic conductivity) at very low carrier density from the finite intercept of the mobility – carrier density curves, and from the observation of strongly localised transport in more depleted samples.

A strong carrier density dependent anisotropy of transport mobility was observed. The highest mobility was always seen for current flow along the $[01\bar{1}]$ direction and the lowest always observed along the [011] direction. As our wafers were aligned with the (100) plane to ±0.1°, we reject misalignment as a source of this anisotropy. This type of anisotropy has also been observed in nominally aligned (100) inverted pseudomorphic InGaAs HEMTs (chapter 5 ref.6) and also on nominally aligned (001) InP based HEMTs, with a pseudomorphic InGaAs channel and AlInAs spacer (chapter 5 ref.7). In

the latter work, the relative difference in mobility between the two directions above increased with the growth temperature and the channel indium content. We report the same conclusion with this particular system, but in addition, show that quantum mobility is independent of direction, indicating that anisotropy is dominated by short-range, large q scattering. In addition, we have shown that the intermediate directions have an intermediate transport mobility.

Scattering and Mobility at Low Temperature After a Bias Cool (sect. 5.4 and 5.8)

The relative effect of the RII scattering could be 'tuned' by the process of cooling under a gate bias. This process altered significantly the occupation correlations established between electrons trapped in the deep-donor centres of the doping layer. The application of a positive bias during cooling enhanced the occupation of such centres, while this was discouraged with a negative bias cooling potential. To varying degrees, all the samples tested with this method displayed changes to both the transport and quantum mobilities of the channel electrons arising from this correlation tuning.

There was a tendency for the transport mobility to change in an approximate step-like manner over a narrow range of bias cool voltage with a saturated low-mobility level at higher negative values. At this saturation, the channel electrons were subjected to the full scattering effect of the unoccupied and hence uncorrelated ionised donors. Saturation was also found in the carrier density of the channel over the same bias cool voltage range, with the corresponding occupational state energy of the donor layer passing through that for deep donors and saturating at a value commensurate with the shallow states.

Although transport mobility is weighted toward large q scattering, the reduction of this mobility measure indicated some influence from smaller q scattering. Quantum mobility was more strongly affected as the remote ionised donors have a strong influence on small q scattering. As with other parameters, quantum mobility was strongly influenced by the degree of electron occupation of the shallow states of the doping layer. Overdoped samples showed significantly higher quantum mobility than those critically doped. To observe the 'true' effect of the uncorrelated donors, these electrons in shallow states were first electrostatically removed.

Application of Matthiessen's rule supported the evidence that the RII scattering in a 5nm spacer sample was consistent with theory, but that the thinnest spacer (2.5nm) samples experienced an additional scattering factor over and above the intrinsic channel scattering mechanism and the RII factors mentioned above.

Scattering and Mobility at Variable Temperature (sects. 5.5, 5.7 and 5.9)

At lower carrier densities in the metallic regime, all samples tested showed a general decrease in mobility or conductivity with decreasing temperature. The magnitudes of the changes are reminiscent of weak localisation and the onset of coherent back scattering. This weak localisation behavour is a precursor to a temperature dependent strong localisation dominant at lower carrier concentrations. In the data for all samples, the zero mobility cut-off shifts to lower carrier density with increasing temperature. This suggests that trapped electrons within the potential fluctuations at high depletion can be thermally activated out of these as the temperature is increased.

A measure of weak localisation was confirmed in both temperature dependent 2-terminal conductivity and 4-terminal magneto-resistance measurements. In the latter, weak localisation was observed provided the temperature (less than about 15K) and carrier density was appropriate. In this case, there was good agreement between the conductivity changes determined independently by varying the magnetic field and the temperature. At very high charge carrier depletion, strong activation was observed at higher temperatures, and large magnitude tunneling currents at low temperatures. No convincingly strong evidence was found in favour of any of the known hopping transport mechanisms, but a larger range of temperatures may be needed to test this exhaustively.

At high carrier density in the metallic regime, parallel conduction and eventually phonon scattering can explain much of the behaviour above about 30K. Magneto-resistance plots for a range of samples suggested the onset of possible parallel conduction (rising B^2 dependency) at a temperature above about 30K. At high carrier densities in samples with the 2.5nm spacers, there was an additional reduction in transport mobility as the temperature fell below 30K to the base temperature. The reasons for this are unknown, though it is somewhat reminiscent of weak localisation behaviour. By room temperature, the mobility had fallen to a value of about a half to a quarter of its low temperature value, depending on spacer thickness, but some elastic scattering factors such as the mobility peaking and the anisotropy persisted.

POSSIBLE FUTURE WORK

Apart from extending the existing techniques used in this project to more of the present samples, the growth and fabrication of new samples, combined with analytical techniques not yet used, would enhance some of the present evidence. Extending both the indium content and the growth temperature range would complement the present data in determining the likely scattering factors both in general terms and for explaining individual features such as anisotropy. Study of transport using even thinner spacers could help to clarify the direct donor scattering mechanism. In addition, this process lacks a complete theoretical treatment. Finally, microscopical techniques such as SEM and TEM on thinned samples should be applied, as many questions can be clarified simply by directly or indirectly viewing the scattering location.

ACKNOWLEDGEMENTS

I am particularly grateful to my supervisor Dr. Andrew R. Long for both giving me the opportunity to commence this undertaking and showing great patience during it. He is due credit for applying most of the theory used. I would also like to thank Prof. John N. Chapman for taking on the role of second supervisor. Much assistance was gratefully received from many colleagues, particularly Dr Elef. Skuras, who patiently demonstrated many of the fabrication procedures, and Brian Milton, who constructed many of the data acquisition algorithms. I would also like to extend gratitude to all those who assisted with mysterious sounding 'private communications', particularly Dr. John H. Davies, who also considered some parts of the theory. I would also like to thank Dr. C. John Emeleus and two undergraduate students – A. Watt and M. O'Leary for providing some of the data for 2 samples.